McGraw-Hill Encyclopedia of Science and Technology

McGraw-Hill Encyclopedia

McGRAW-HILL BOOK COMPANY INC.

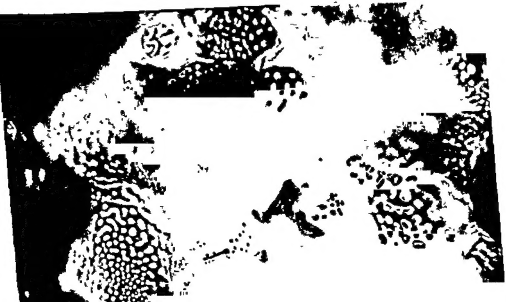
NEW YORK CHICAGO SAN FRANCISCO DALLAS TORONTO LONDON

# of Science and Technology

AN INTERNATIONAL REFERENCE WORK

IN FIFTEEN VOLUMES INCLUDING AN INDEX

VOLUME 14 TOW-ZYT



(LEFT) Transgranular cleavage patterns observed on the fracture surface of a piece of high-density beryllium oxide (beryllia) *(photograph by W. C Coons, Curtiss-Wright Corp).* (RIGHT) Bubble structure in grain boundary areas of hot-pressed beryllia. Beryllia is transparent when viewed through an oil-immersion objective. Actual focus 0.008 mm below the polished surface of the specimen *(photograph by W. C. Coons, Curtiss-Wright Corp)*

Library of Congress Catalog Card Number: 60-11000

Suggestions to the Readers

The basic plan of the Encyclopedia is explained here in order to facilitate its use

The subject matter of the various disciplines or branches of science and technology is organized systematically a general article provides a broad survey of the field, and a number of separate articles, alphabetically arranged, cover its main subdivisions and more specific aspects.

Cross references guide the reader from the general articles to the other articles into which the subject is subdivided, and from these to articles on more highly specialized phases of the subject The cross references—there are about 50,000 of them—are printed in small capital letters so that they can be easily recognized By means of the cross references a reader may find his way from ELECTRICAL ENGINEERING, through ELECTRONICS and VACUUM TUBE, to ELECTRON MOTION IN VACUUM or ELECTRON EMISSION Or, following another line of cross references, the reader would be led to ELECTRIC POWER SYSTEMS, TRANSMISSION LINES, ELECTROMAGNETIC WAVE, and so on

In general each article begins with a definition of the title that states its scope and coverage Usually, only the scientific or technological sense is discussed. Most of the articles, after this statement, go on to increasingly complex and detailed considerations A reader thus needs to proceed only as far as his inclinations and requirements dictate.

The Index, Volume 15, should be consulted to locate the discussion of topics covered in the Encyclopedia but not given in separate entries

Every phylum, class, and order in the plant and animal kingdoms is allotted a separate article Many of the more common families, genera, and species are covered either in one of the order articles or in a separate article under its own scientific or common name

The adjectives electric and electrical are used in the following senses Electric—containing, producing, arising from, actuated by or carrying electricity, or capable of doing so; as, for instance, electric generator, electric motor, electric wiring. Electrical—related to, pertaining to, or associated with electricity, but not having its properties or characteristics; as, for example, electrical code, electrical engineering

Words used as titles are, wherever possible, given in the singular to permit a consistent alphabetic arrangement. Titles are alphabetized by word and not by letter, for example,

Earth sciences
Earth tides
Earthmover
Earthquake

A word used as a noun precedes the same word used adjectivally, thus

Mercury (element)
Mercury (planet)
Mercury battery

or

Circuit, electronic
Circuit breaker

Hyphened terms are alphabetized as single words. for example,

Animal virus
Animal-feed composition

Most of the longer articles contain bibliographies citing useful sources of further information For additional bibliographical citations, the reader should refer to related articles (as indicated by the cross references in the article) Bibliographies are placed at the ends of articles or sometimes at the ends of major sections in long articles

A list of initials and names of the contributors to the Encyclopedia is to be found in Volume 15 This list will permit quick identification of a contributor's initials after an article Immediately following this list is a second list of encyclopedia contributors with their affiliations and the titles of articles each has written for the Encyclopedia

McGraw-Hill Encyclopedia of Science and Technology

Tower

A concrete, metal, or timber structure that is relatively high for its length and width. Towers are constructed to support transmission lines, radio and television antennas, and for many other purposes, including rocket and missile launching.

Transmission towers. These are rectangular in plan and are not steadied by guy wires. A transmission tower is subjected to a number of forces: its own weight, the pull of the cables at the top of the tower, the weight of wind and ice on the cable, and the effect of wind on the tower itself. Torsional forces on a tower caused by breakage of the cables in one span on one side of the tower must also be considered.

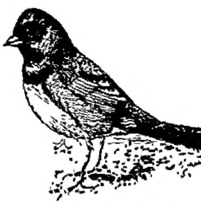
Radio and television towers. These are either guyed or free-standing. Free-standing towers are usually rectangular in plan. In addition to their own weight, free standing towers support the weight of the antenna and accessories, and the weight of ice, unless a deicing circuit is installed. Wind forces must also be carefully considered.

Guyed towers are usually triangular in plan with the main structural members, or legs, at the vertices of the triangle. The legs are usually solid round steel bars. All members are galvanized and primed before erection. Field connections are made with galvanized bolts and lock type nuts. The original design usually provides for increasing the height of the tower, when this is permitted by the Federal Communications Commission.

In 1959 construction was underway on a structure in Moscow that was to be the world's tallest. The structure, a 1667 ft television and observation tower of reinforced concrete, was to be 23 ft in diameter at the top and 213 ft across at the bottom. See TRESTLE. [C.N.C.]

Towhee

Any of seven species of moderately large American sparrows of the genera *Pipilo* and *Chlorura*, four of which occur in the United States. Best known is the redeyed towhee, *Pipilo erythrophthalmus*, also called the spotted towhee, joree, and chewink. All the common names refer to its loud, oft-repeated call. The males are flashy black, chestnut, and white, the female is brown where the male is black. They live in dense, brushy cover, where they scratch in the leaves for food. This is the only towhee found in the eastern United States, it occurs throughout the country. Several distinct subspecies are recognized, such as the white- and red-eyed races. See PASSERIFORMES; SPARROW. [J.D.B.]



The towhee, *Pipilo erythrophthalmus*; length to 8¼ in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Towing tank

A tank of water used to determine the hydrodynamic performance of water-borne bodies such as ships and submarines, as well as torpedoes and other underwater forms. In the narrow sense, towing tanks are considered to be experimental facilities used to measure the forces, such as drag, on ship models and in turn to predict the performance of the full-scale prototype. In general, towing tanks are rectangular in plan form with a uniform cross section. Differ...

from rectan...

dimension...

from about 4 to 22 ft in depth, and from under 100 to almost 3000 ft in length; the size of the model varies in length from 4 to 30 ft.

Towing methods. The principle measurements made in a towing tank are force measurements,

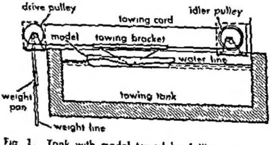


Fig. 1. Tank with model towed by falling weight.

particularly drag or resistance of a towed ship model or other body. One of two principal systems for towing a model is used in most towing tanks. The simpler system consists of a gravity dynamometer and an endless cable attached to the model. A weight provides a constant towing force (Fig. 1). The time to traverse a fixed distance is measured when the model reaches a constant speed, thus establishing the speed-resistance relationship for the model. This dynamometer is simple, capable of high accuracy, but is limited to the measurement of the drag force of water-borne bodies. It is used in the smaller towing tanks where the models are generally under 6 ft in length.

In larger towing tanks the model is towed by a towing carriage mounted on rails at the side of the towing tank or suspended from an overhead track system (Fig. 2). Speed can be controlled and measured precisely on these carriages Most carriages are equipped with a drag dynamometer as a permanent component (Fig 3).

The dynamometer girder, mounted on the carriage frame, carries a long horizontal floating beam in pendulum fashion on two pairs of vertical arms terminating in flexible springs. A counterweight at the upper end of a vertical swinging arm mounted on the girder and attached to the floating beam maintains the beam in equilibrium at any position between the limit stops The model resistance is transmitted as a horizontal force through the upper flexible link L_1 to the lower arm of the T-shaped balance, where it is balanced by the weight W. When the model resistance is not equal exactly to a unit weight W, the remainder is taken up (or applied) by the resiliency of the whole group of flexible spring supports; the exact amount of this auxiliary load or force is recorded on the drum by the link L_2 and the recording arm shown.

In modern installations when the drag forces on submerged bodies must be measured it is usually accomplished by electrically measuring the strains in a flexure. The strain gage dynamometer is located at the point where the model is attached to the supporting strut.

Law of similitude. Modern towing tank technology was established by William Froude in the 1870s, when he discovered the law of similitude for phenomena in which gravity is the predominating factor and established one of the essential principles of hydrodynamics for comparing model phenomena with the actual ship. There are three principal forces involved for a body moving through the water: inertia, gravity, and viscosity. The law of similitude requires that the ratio between inertia force and gravity force be the same for both the model and the prototype, and that the ratio between inertia force and viscous force be the same as well. *See* Froude number; Reynolds number

Froude's law requires

$$\frac{V}{v} = \sqrt{\frac{L}{l}} = \sqrt{\lambda} \qquad (1)$$

where V and v represent the velocity of the prototype and model, respectively; L and l represent a characteristic dimension such as length of prototype and model, respectively, and λ represents the linear ratio of prototype to model.

Reynolds' law requires

$$\frac{VL}{\nu_s} = \frac{vl}{\nu_m} \qquad (2)$$

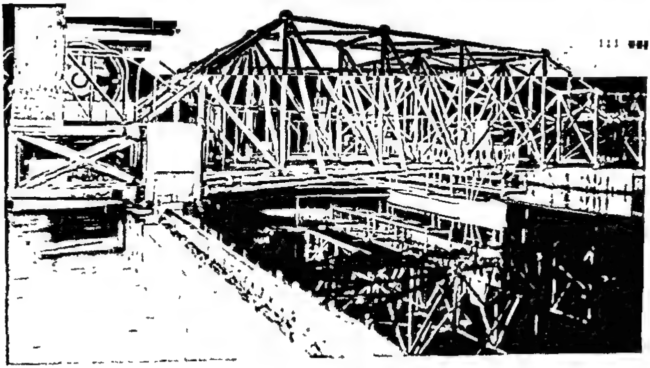


Fig. 2. Model in David Taylor Model Basin is towed by carriage

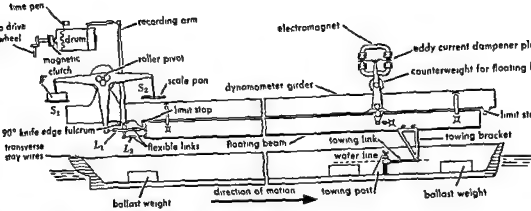


Fig 3 Schematic arrangement of towing dynamometer

where ν_s and ν_m are the kinematic viscosity of the fluid for prototype and model, respectively. Because model-towing tanks use fresh water and ships usually operate in sea water, ν_s and ν_m are essentially in the same order of magnitude. Thus, Froude's and Reynolds' laws require a different velocity relationship between the prototype and the model The law of similitude cannot be satisfied in the strict sense for test of ship model in water

Froude overcame this difficulty by dividing the [illegible] primarily the effect of viscosity and is thus governed by Reynolds' law Residuary resistance is due mainly to gravitational effect and thus is governed by Froude's law Based on this distinction, Froude developed the technique of predicting ship resistance from the resistance test of a scaled model, a technique used to this day. It can be outlined as follows A model geometrically similar to the prototype is made and the drag or resistance measured in a towing tank at the corresponding speeds as expressed in Eq (1) Viscous or frictional resistance r_F of the model is calculated by assuming the resistance to be the same as that of a smooth flat plate of comparable area and length Residuary resistance r_R, which is largely a gravitational effect, is then found by subtracting the frictional from the total measured resistance r_T

$$r_R = r_T - r_F \tag{3}$$

Residuary resistance of the ship R_R is calculated by the law of comparison at corresponding speeds,

$$R_R = r_R \times \lambda^3 \tag{4}$$

The frictional resistance of the ship R_F is calculated on the same assumptions used in calculating the viscous resistance of the model. The total resistance of the ship is finally found by adding the residuary and frictional resistance

$$R_T = R_F + R_R \tag{5}$$

Besides the resistance test, there are two other important tests: the propeller open-water test and the self-propulsion test. From the measuring of the rpm, torque, and thrust of the model propeller, the shaft horsepower required to drive the prototype at designed speed can be predicted

Force measurements. Experimentation in modern towing tanks has been extended far beyond predicting the resistance and power of a moving body *in still water.* Towing tanks are used to measure any combination of forces and moments upon a water-borne or submerged body under steady-state conditions. A few of the many possible types of tests involving force measurements in a towing tank are (1) lift and drag of a planing surface; (2) lift, drag, and pitch moment of a submerged hydrofoil, (3) forces and moments on a submerged body towed at an angle of attack to obtain the static coefficients of a body for equations of motion; and (4) the turning moment on a ship's rudder.

Pressure and velocity measurement. A series of experiments has evolved from the measurement of pressures and velocities The pressure measurements may be integrated over the surface to determine the force on the surface. Typical tests are the measurement of the pressure distribution on propeller blades, the duct of a ducted propeller, and the appendages of ship models. Various velocity-measuring devices have been developed, the most common of which is the pitot tube. It is used to measure the velocity field at various locations around a model's hull, the most common of which is the plane in which the ship's propeller operates.

Lines of flow. A towing tank is also used to determine the lines of flow over portions of a hull; thus, appendages can be installed so that they will have minimum resistance and avoid or minimize the creation of cavitation For surface ships, the conditions of comparable speed are maintained between model and prototype Flow patterns on the hull are usually established by the emission of dyes upon a color sensitive paint such as hydrogen sulfide upon a white lead based paint. Flow patterns outside the boundary layer may be established by vanes or flags, which are free to pivot and

orient themselves to the direction of flow. Typical tests are (1) establishment of the location of bilge keels; (2) orientation of shaft struts; and (3) determination of wave profile.

Wave experiments. In recent years towing tanks have been provided with wavemakers, which have extended the range of experimentation to the study of ship performance in head and following seas. At one end of the towing tank a wavemaker is installed which can generate waves of a more or less uniform profile with a predetermined height and length. At the other end of the tank there is installed a wave absorption beach. Experiments are conducted on ship models at corresponding speeds in various head and following sea conditions, recognizing that the waves generated are much more regular than those encountered in the ocean (*see* OCEAN WAVES). In addition to maintaining geometric similarity of the model and its prototype, the dynamic similarity of the system must also be maintained. Measurements are made primarily of the motions of the body, particularly the rotational motions, pitch and roll, and the translational motions, heave (vertical) and surge (longitudinal). The most recent techniques involve free-running models with sensitive accelerometers. The carriage is used to carry the recording equipment and provide power to the model through flexible cables which do not exert any restraining force on the model. In some instances in connection with the measurement of motions various components of force or moment may also be measured.

Nonsteady-state experiments. With the advent of electronics the study of unsteady hydrodynamic phenomena has become increasingly important in towing tanks. Such fluctuating forces or pressures are measured as (1) pressure on a model's hull from propeller blades passing in close proximity, (2) vibratory forces produced by a propeller operating in the variable velocity field behind a model; (3) route stability characteristics of a model from alternate course variations, (4) forces and moments on a submerged body undergoing pure pitching and heaving motion. From the last measurement the coefficients for the equation of motion are obtained for a specific submarine design. This makes possible, through the use of the analog computer, the calculation of a submarine's motion without further experimentation and under a variety of conditions difficult to achieve in a towing tank.

New tanks. The towing tanks and the associated test equipment described have spawned a number of special purpose test facilities which are modifications of the tanks described above. The most important of these are the turning, rotating arm, and seakeeping tanks or basins. As its name implies, the turning tank is used to conduct experiments on the turning characteristics of ship models. Again the corresponding speed must be maintained to obtain similarity of phenomena. The rotating arm is a specialized facility used to find the so-called rotary derivatives of bodies such as submarines and torpedoes. From the measurement of the forces and moments on a body traversing a curved path, the rotary components of the equations of motion for a particular body are derived. The seakeeping basin is designed to study the motions and forces on water-borne bodies under various sea conditions. It may also be used to generate more complex seas than can be created in a towing tank with a single wavemaker. *See* WATER TUNNEL; WAVE MOTION IN LIQUIDS; *see also* SHIP PROPULSION. [J.B.H.]

Bibliography: H. E. Saunders, *Hydrodynamics in Ship Design*, vol. 1, 1957.

Townsend discharge

A particular part of the voltage-current characteristic curve for a gaseous discharge device named for J. S. Townsend, who studied it about 1900. It is that part for low current where the discharge cannot be maintained by the field alone. Thus, if the agents producing the initial ionization were removed, conduction would cease.

In the lower end of this region, conduction is accomplished only by charges produced by external agents. As the electric field is increased, secondary ionization and more efficient collection of the primary ionization causes an increase in the current. After further increase in the field, the end of the Townsend region is reached. Any further increase in field causes a transition into a region where the discharge may be maintained by the field alone, whether it be glow, brush, or arc. *See* DARK CURRENT; ELECTRICAL CONDUCTION IN GASES; GLOW DISCHARGE. [G.H.MI.]

Toxemia

A generalized state in which the blood contains toxins, usually of bacterial origin. Toxins are soluble poisons of two general types, the exotoxins secreted by bacteria into surrounding fluids, and endotoxins retained in the organisms until set free by its disintegration. Most toxins are probably enzymes and, after a latent period, cause tissue damage which in turn produces symptoms. These include malaise, easy fatigability, generalized aching, and fever. Certain toxins are rather specific in their action, as in the case of the neurotoxins of food poisoning. Others may produce injury to red blood cells, causing jaundice, or may interfere with clotting, producing hemorrhages. Organ injury may result in congestion, albuminuria, hemorrhage, and fluid accumulation in body cavities. *See* BLOOD, ENZYME, FOOD POISONING, BACTERIAL; TOXIN, BACTERIAL.

A special group of diseases, the toxemias of pregnancy, occur in 6–7% of women during or immediately following pregnancy. Hypertension, edema, and proteinuria may be present and in severe cases eclampsia follows with convulsions and coma. These toxemias constitute one of the three major causes of maternal death, the other two being hemorrhage and infection; 30,000 infants of toxemic mothers are stillborn or die at birth each year in the United States. Although the causes are

obscure, early diagnosis and treatment could make almost every case preventable. *See* PREGNANCY.

[E.G.ST.]

Toxicology

The study of poisons and their effects, mechanisms of action, and methods of treatment. The term poison is difficult to define because any material, if used in large enough amounts or administered in certain ways, will produce harmful effects on some structure or function of the body. Other factors which must be considered include an individual's general state of health, age, sex, and whether previous exposure to an agent has increased his tolerance to it or has caused a cumulative effect which renders him more susceptible to an added dose.

For all practical purposes, poisons are substances which cause tissue damage or malfunction of a potentially serious degree when given to an average individual in small amounts. This amount is usually considered to be less than 50 g. In addition, poisons must exert their effects through chemical or physicochemical mechanisms. The above definition eliminates hypersensitivity reactions which do not occur in the average person, and also excludes agents that cause purely physical damage. *See* HYPERSENSITIVITY.

Categories of poison. Different poisons produce their damage in various ways. Four broad categories may be enumerated.

1. Agents which act so rapidly that no perceptible direct effects are seen; death occurs swiftly as a result of anoxia, usually following circulatory collapse. Carbon monoxide and cyanide are asphyxiant gases of this category. The alcohols, ethers, and other hydrocarbons which are central nervous system depressants are also included. In each case, a lesser amount of the same substance may not be rapidly fatal and then direct chemical effects may be seen in susceptible tissues. *See* CENTRAL NERVOUS SYSTEM.

2. Many agents produce damage at point of contact or entry into the body. Such locally destructive poisons include such corrosives as acids or alkalies, and also irritant gases, volatile oils, and aconite.

3. Systemic poisons exert their principal effects after they are absorbed in the body. They may also cause some local irritation at the point of entry to the body and therefore need [illegible] system, or blood forming tissues. Agents of this group include arsenic which in acute cases causes fatal blood vessel injury, and other heavy metals that inhibit enzymic activity at cellular levels. Lead, mercury salts, thallium, cadmium, phosphorus, manganese, and chromium each cause fairly specific [illegible]

4. [illegible] some [illegible] damage is seen at the point of entry. These include the blood destroying (hemolytic) poisons such as nitrobenzene and arsine, and the central nervous system depressants, such as hydrogen sulfide and the war gases of the nerve-gas type.

Another point of view in classification of poisons considers the source of each noxious substance, even though chemical characteristics and effects may be quite different in each group. The cate- [illegible] industrial and occupational [illegible], and animal [illegible] such a classification is useful in both preventive and diagnostic medicine.

Other descriptive terms used in reference to poisons reflect the mechanism of action or the principal tissues involved. Examples include corrosives, irritants, asphyxiants, enzyme inhibitors, neurotoxins, and hemotoxins.

Statistical data. Statistical studies of poisonings vary a great deal in their conclusions, chiefly because of differences in geographic location, occupational and social status of those involved, and the availability of good laboratory facilities for toxicological analysis.

About 55% of deaths from poison are accidental, 45% are suicidal, and fewer than 1% are homocidal. Children are most often involved in accidental injury following ingestion or exposure to various substances. Of the deaths, 55% follow aspirin ingestion, 25% result from damage produced by other drugs, and the remainder include the more common cleaning agents, cosmetics, kerosine, lye, paint products, pesticides, and animal venoms.

Industrial and agricultural workers are frequently exposed to poisons through accidents or by improper use of safety measures. Chlorinated hydrocarbons, nitro and amino aromatics, benzene, carbon monoxide, chromium, mercuric salts, methyl alcohol, hydrogen sulfide, arsenic, DDT, and other insecticides account for a large proportion of cases and not infrequent fatalities.

Carbon monoxide and overdoses of barbiturates are probably the most commonly used chemicals in suicides, although many other materials are used, despite the agonizing results known to be produced by some, such as lye.

Ethyl alcohol, carbon monoxide, acids, alkalies, cyanide compounds, mercury preparations, arsenicals, sedatives, and benzene are probably the most commonly encountered agents in poisonings, if one excludes the special categories of insect and snake bites, food poisoning, and certain plants, such as inedible mushrooms, that contain harmful materials.

History. The history of poisons is as old as man. Even today some of the most primitive peoples of the world possess substances whose development is shrouded in mysticism and superstition. Often the ingredients have remained impervious to modern analysis; occasionally one of these poisons is purified, standardized, and introduced into accepted medical use. For example, curare originated as arrow poison for the South American Indians of

orient themselves to the direction of flow. Typical tests are (1) establishment of the location of bilge keels; (2) orientation of shaft struts; and (3) determination of wave profile.

Wave experiments. In recent years towing tanks have been provided with wavemakers, which have extended the range of experimentation to the study of ship performance in head and following seas. At one end of the towing tank a wavemaker is installed which can generate waves of a more or less uniform profile with a predetermined height and length. At the other end of the tank there is installed a wave absorption beach. Experiments are conducted on ship models at corresponding speeds in various head and following sea conditions, recognizing that the waves generated are much more regular than those encountered in the ocean (*see* OCEAN WAVES). In addition to maintaining geometric similarity of the model and its prototype, the dynamic similarity of the system must also be maintained. Measurements are made primarily of the motions of the body, particularly the rotational motions, pitch and roll, and the translational motions, heave (vertical) and surge (longitudinal). The most recent techniques involve free-running models with sensitive accelerometers. The carriage is used to carry the recording equipment and provide power to the model through flexible cables which do not exert any restraining force on the model. In some instances in connection with the measurement of motions various components of force or moment may also be measured.

Nonsteady-state experiments. With the advent of electronics the study of unsteady hydrodynamic phenomena has become increasingly important in towing tanks. Such fluctuating forces or pressures are measured as (1) pressure on a model's hull from propeller blades passing in close proximity; (2) vibratory forces produced by a propeller operating in the variable velocity field behind a model; (3) route stability characteristics of a model from alternate course variations; (4) forces and moments on a submerged body undergoing pure pitching and heaving motion. From the last measurement the coefficients for the equation of motion are obtained for a specific submarine design. This makes possible, through the use of the analog computer, the calculation of a submarine's motion without further experimentation and under a variety of conditions difficult to achieve in a towing tank.

New tanks. The towing tanks and the associated test equipment described have spawned a number of special purpose test facilities which are modifications of the tanks described above. The most important of these are the turning, rotating arm, and seakeeping tanks or basins. As its name implies, the turning tank is used to conduct experiments on the turning characteristics of ship models. Again the corresponding speed must be maintained to obtain similarity of phenomena. The rotating arm is a specialized facility used to find the so-called rotary derivatives of bodies such as submarines and torpedoes. From the measurement of the forces and moments on a body traversing a curved path, the rotary components of the equations of motion for a particular body are derived. The seakeeping basin is designed to study the motions and forces on water-borne bodies under various sea conditions. It may also be used to generate more complex seas than can be created in a towing tank with a single wavemaker. *See* WATER TUNNEL; WAVE MOTION IN LIQUIDS; *see also* SHIP PROPULSION. [J.B.HA.]

Bibliography: H. E. Saunders, *Hydrodynamics in Ship Design*, vol. 1, 1957.

Townsend discharge

A particular part of the voltage-current characteristic curve for a gaseous discharge device named for J. S. Townsend, who studied it about 1900. It is that part for low current where the discharge cannot be maintained by the field alone. Thus, if the agents producing the initial ionization were removed, conduction would cease.

In the lower end of this region, conduction is accomplished only by charges produced by external agents. As the electric field is increased, secondary ionization and more efficient collection of the primary ionization causes an increase in the current. After further increase in the field, the end of the Townsend region is reached. Any further increase in field causes a transition into a region where the discharge may be maintained by the field alone, whether it be glow, brush, or arc. *See* DARK CURRENT, ELECTRICAL CONDUCTION IN GASES, GLOW DISCHARGE [G.H.MI.]

Toxemia

A generalized state in which the blood contains toxins, usually of bacterial origin. Toxins are soluble poisons of two general types, the exotoxins secreted by bacteria into surrounding fluids, and endotoxins retained in the organisms until set free by its disintegration. Most toxins are probably enzymes and, after a latent period, cause tissue damage which in turn produces symptoms. These include malaise, easy fatigability, generalized aching, and fever. Certain toxins are rather specific in their action, as in the case of the neurotoxins of food poisoning. Others may produce injury to red blood cells, causing jaundice, or may interfere with clotting, producing hemorrhages. Organ injury may result in congestion, albuminuria, hemorrhage, and fluid accumulation in body cavities. *See* BLOOD; ENZYME; FOOD POISONING, BACTERIAL; TOXIN, BACTERIAL.

A special group of diseases, the toxemias of pregnancy, occur in 6–7% of women during or immediately following pregnancy. Hypertension, edema, and proteinuria may be present and in severe cases eclampsia follows with convulsions and coma. These toxemias constitute one of the three major causes of maternal death, the other two being hemorrhage and infection; 30,000 infants of toxemic mothers are stillborn or die at birth each year in the United States. Although the causes are

Toxoids are detoxified, antigenic toxins. Formaldehyde and a few other agents will destroy the toxicity of the typical exotoxins without impairing their antigenicity, that is, their ability to evoke antitoxins. Such toxoids are extensively used in vaccines, diphtheria and tetanus toxoids being important examples. *See* ANTIGEN; ANTITOXIN; LETHAL DOSE 50; TOXIN-ANTITOXIN REACTION. [H.P.T.]

Bibliography: W. E. Van Heyningen, *Bacterial Toxins*, 1951

Toxin-antitoxin reaction

A term used in serology to denote the combination of a toxic antigen with its corresponding antitoxin. If the antitoxin is derived from any species other than the horse, precipitation occurs over a wide range of reactant ratios, 20× or more, as in other antigen-antibody reactions. With horse antitoxin, flocculation occurs only if toxin and antitoxin are near equivalence, a twofold excess of either reactant giving soluble complexes. In most instances, the reaction results in partial or complete neutralization of the toxic activity of the antigen. *See* ANTIGEN; ANTITOXIN; NEUTRALIZATION REACTION (ANTIBODY).

The Danysz reaction occurs when an exact equivalence of toxin is added to antitoxin, not in one portion, but in successive increments. The mixture will remain somewhat toxic, although it becomes neutral within a few days. This is the result of the variable combining proportions of antigen and antibody, which results in a binding of excess antibody by the initial increments of toxin giving soluble molecular complexes composed of one molecule of toxin and eight molecules of antibody (TA_8). As a result some of the subsequent increments of toxin remain free. In time, the system rearranges to give the neutral equivalence composition, one molecule of toxin and two molecules of antitoxin (TA_2). *See* ANTIBODY; SEROLOGY [H.P.T.]

Bibliography: M. Cohn and A. M. Pappenheimer Jr., A quantitative study of the diphtheria toxin-antitoxin reaction in the sera of various species, including man, *J. Immunol.*, 63:291–312, 1949.

Toxoplasmosis

A disease of man and animals produced by a protozoin, *Toxoplasma gondii*. Although infection with this microorganism is usually not apparent or even recognized in adults, in children it may take the form of various syndromes in which the viscera and central nervous system are involved. The discovery that *Toxoplasma* is a common parasite of man and animals, of world-wide distribution and high prevalence in many human populations, for example, the United States, has been one of the important findings made concerning infectious diseases since World War II. *See* HAEMOSPORIDIIDA.

Humans become infected in two ways, congenitally and from the external environment after birth. These external sources appear to have become significant by 5 years of age. From then on, infection rates in the general population of the United States increase with age, until a maximum is reached between 30 and 50 years at which time 30–60% of normal adults, according to the area surveyed, give serological evidence of infection, past or present, with *Toxoplasma*. *See* SEROLOGY.

Infants may be born with Toxoplasma contracted from their mothers. This congenital form is often fatal, the brunt of the toxoplasma attack being borne by the brain, spinal cord, and eyes. The mothers of these children are almost invariably healthy, and probably most infections in adults are well tolerated. When adult disease does occur it can be mild or fatal; the organs and tissues involved vary and are often multiple, including lymph nodes, endocrine glands, heart, liver, lungs, brain, and skin.

Although *Toxoplasma* possesses some highly effective mechanisms for infecting man, these sources have not all been discovered. The eating of undercooked pork is in all probability one such source. Since the epidemiology is not clearly understood, rational methods to effect complete control of the infection are not now available.

Diagnosis depends upon isolation of *Toxoplasma* or upon immunological procedures which detect specific antibodies in the blood serum. Present treatment is only moderately successful; useful compounds include the sulfonamides and the pyrimidines. *See* ANTIBODY; IMMUNITY; PROTOZOA; SULFA DRUGS. [D.W.]

Trace analysis

The determination of small amounts of material present in low concentration in fairly large samples. This contrasts with microanalysis in which a major constituent in a very small sample is determined. For example, in drinking water the calcium content may be as low as a few milligrams per liter, and the sample used may be 50 milliliters. Optical, electrical, or radioactivity methods are usually used. Trace analysis is most important in cases where the concentration of impurity is critical, and where small amounts of an added substance have a major effect on the properties of the total sample as in semiconductors. *See* ANALYTICAL CHEMISTRY. [K.G.S.]

Tracer, radioactive

A radioactive isotope which, when injected into a chemically similar substance, or artificially attached to a biological or physical system, can be traced by radiation detection devices. Many problems in biology, medicine, and industrial engineering not amenable to other approaches can be solved by the use of these radioactive tracers. *See* RADIOACTIVITY; RADIOACTIVITY (APPLICATIONS).

Uses in biology [illegible] by following the radiation from the isotope. This fundamental application of radioactive tracers has been used to study the nocturnal behavior of bats. One investigator attached small sources of cobalt-60 to the legs of the bats he was studying. A radiation detector connected to an au-

the Amazon and Orinoco regions. Although a synthetic compound now has been developed, the action is similar, but instead of using amounts that cause fatal muscle paralysis, the physician uses lesser quantities to relax the muscular system.

Other medicinals which have been handed to civilization from witch doctors and ancient priesthoods include strychnine, opium, caffeine, strophanthin, cocaine, atropine, digitalis, and ergotamine. Still more compounds, usually prepared from certain plants, have been used from time to time to remove an unwanted person, or more commonly, have been used by large numbers of a population as partial poisons to produce certain effects. The most notable of these are opium, hashish, caffeine, nicotine, marijuana, heroin, betel, and fly agaric. The deadly nightshade, henbane, mandrake root, and the thorn apple are examples of plants which have been used and misused because of their "magical" properties, which permit the mind to experience fantastic states or be plunged into a narcolepsy, impervious to pain, fatigue, or sorrow.

Modern man has added innumerable chemicals capable of potential injury, even though most have been isolated or synthesized for beneficial reasons Modern chemical and pharmaceutical processes have developed hundreds of new products which are used daily. Many are poisons whose usefulness outweighs the occasional instances of sickness or death they cause.

Treatment for poison. Treatment depends on the nature of the poison, the amount, and the route of administration. First aid measures are directed toward prevention of further absorption by the body, removal of poisons when possible, neutralization by general or specific antidotes, and support of failing body functions. For inhaled poisons, fresh air, oxygen, artificial respiration, and maintenance of body temperature may be employed as needed until trained aid is available. For ingested poisons, a specific antidote is given, if known, or the universal antidote, if the agent is unknown The universal antidote consists of two parts powdered charcoal, one part tannic acid, and one part magnesium oxide. A household substitute can be made from two pieces of burned toast, one part of strong tea, and one part of milk of magnesia. Other materials may be given to dilute the poison and slow absorption, such as milk, beaten eggs, or a suspension of flour, starch, or eggs in water. Vomiting should be induced in conscious patients unless a corrosive has been ingested, in which case regurgitation may cause further damage.

For skin and eye contaminations, water is used to flush off the poison. No chemical antidotes should be employed by the untrained.

Injected poisons from snake bites or drug overdosage should be immobilized as much as possible at the site of entry. The patient is kept still, a tourniquet is placed above the site, an ice pack is applied to reduce spread, and incision with suction may be used. Polyvalent venom antiserum should be available in areas with a high incidence of snake bite

In the United States, the establishment of Poison Control Centers has increased rapidly in the past few years. These are set up by qualified personnel and are usually located in or near a cooperating hospital. Their purpose is to be available for disseminating specific information on the toxicity, ingredients, and treatments of poisonings by both general and trade name products. First aid information is given to individuals who telephone, together with information on availability of medical treatment. Physicians may call to obtain specific information for diagnosis and treatment. Most of these centers also have treatment facilities available for local cases. In addition, information on epidemiology, treatment, diagnostics, and mechanisms of action of new poisons, as well as the results of research studies, are passed on to other centers and interested persons through a national clearing house. [F G.ST.]

Toxin, bacterial

Many bacteria produce substances that have a pronounced pharmacological effect, a few such toxins being among the most lethal chemicals known. The typical exotoxins are readily separated from the bacterial cells, are relatively heat-labile, and are easily converted to toxoids Some examples are shown in the table

Potency of some bacterial toxins and an alkaloid

Toxin of	Potency*	Host species tested
Corynebacterium diphtheriae (causative agent of diphtheria)	3 5 3,500	Mouse, guinea pig
Clostridium botulinum, type A (agent of botulism)	620,000 1,200,000	Mouse, guinea pig
Clostridium tetani (agent of tetanus)	200,000 1,200,000	Mouse, guinea pig
Atropine (representative alkaloid for comparison)	0 03	Cat

* Expressed as the number of lethal $dose_{50}$ (LD_{50}) units per milligram toxin

Since a number of test species are shown in the table the potencies are further corrected to a common basis of 1 kg animal weight. Thus, in 1 mg diphtheria toxin, there is 3 5 times more than is sufficient to kill 50% of the individuals in a group of mice whose weights total 1 kg Even when the animal weights are taken into account, some species may be hundreds of times more susceptible to a given toxin than are other species Indication of the test species is therefore important

These toxins are largely responsible for the characteristic effects of the diseases with which they are associated, and antitoxins to them will neutralize their effects, if given in time. Gram negative (dysentery and typhoid) bacteria contain endotoxins that are less-readily separable from the cells, are more heat stable, are not inactivated by formaldehyde, and are less active pharmacologically. Thus the LD_{50} for mice is about 20 times that of diphtheria toxin.

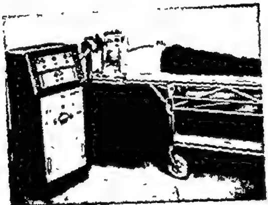


Fig. 2 Brain tumor scanner. (Baird-Atomic Co.)

ing some of the carbon within the ring into carbon-14. This is done by placing the ring in an intense neutron flux. The piston ring is then installed in the engine and, after a period of use, samples of the engine oil are removed and tested for carbon-14 content. In this way amounts of wear much smaller than could be found by weight measurements and other techniques can be determined

On a larger scale, the entire operation of an industrial plant can be tested and controlled by tracer techniques In one industrial survey sulfur-35, prepared in an appropriate chemical form, was introduced into a batch of coke entering an iron-smelting plant. The fate of the sulfur within the plant and its presence in the iron were studied by taking samples and measuring the iron's sulfur-35 content It is entirely possible that similar techniques could be used in the routine operation and monitoring of many industrial processes Such applications have been restricted in the past chiefly because of the difficulty of hazard control, but with adequate safeguards and more sensitive detecting devices, such procedures may soon become commonplace [G.L.B.]

Bibliography See RADIOACTIVITY (APPLICATIONS).

Tracheophyta

Plants having vascular systems consisting of specialized water- and food-conducting tissues Tracheophyta a phylum of the subkingdom Embryophyta, contains over five-sevenths of all known plants or about 260 000 species, distributed throughout the world This phylum includes club mosses, horsetails, ferns conifers, and flowering plants Despite variation in body form from tiny herbs to gigantic trees, all members have certain characteristics in common They are fundamentally land plants although some, such as water ferns and water lilies, have gone back to an aquatic habitat Each plant has xylem (water-conducting tissue) and phloem (food-conducting tissue). All have tracheids or such phylogenetic derivatives as vessel elements and wood fibers.

Reproduction. An alternation of generations is thought to occur in all members The sporophyte (spore-producing generation) is large usually ... can make its ... eration, the ... smaller, less

Origin. The tracheophytes are believed to have evolved from Chlorophyta (green algae) which invaded the land from the sea (see CHLOROPHYTA). It has been suggested that at one stage in the evolutionary process, thalluslike species (structurally simple and frequently prostrate plants) differentiated forming roots, stems, and leaves.

Economic importance. Tracheophyta are of great economic importance. Products obtained from these plants include grain and forage, lumber, cork, textile fibers, cordage, oils, gums, resins, rubber, spices, perfumes, drugs, beverages, dyes, tanning materials, waxes, and paper

The phylum Tracheophyta is composed of four subphyla: Psilopsida, Lycopsida, Sphenopsida, and Pteropsida. *See* LYCOPSIDA; PSILOPSIDA; PTEROPSIDA; SPHENOPSIDA; *see also* PLANT KINGDOM [P.D.S.]

Bibliography: See EMBRYOPHYTA.

Trachoma

An infectious disease of the conjunctiva and cornea of eye caused by a virus of large size which resembles the psittacosis virus and the lymphogranuloma venereum virus. This serious granular conjunctivitis is of wide distribution Cytoplasmic inclusions in the conjunctival cells consist of masses of particles resembling the elementary bodies of the psittacosis-lymphogranuloma venereum group. They stain reddish purple with Giemsa and blue with Castaneda's stain. The particles are held together by inclusion material which stains brown with iodine because of its glycogen content.

The disease is transmitted from eye to eye by fingers or fomites, possibly flies. It is usually acquired early in life by the child from the mother. There is inflammatory swelling of the upper eyelids and granulation of the tarsal conjunctiva may spread to the cornea. In the chronic stage, scarring causes deformity of the eyelids and vascularization or pannus of the cornea Early oral treatment with sulfonamides or topically administered tetracyclines in ointments have reduced the incidence of the chronic stage in the Indians of the United States. *See* LYMPHOGRANULOMA-PSITTACOSIS GROUP [K.F.M.E.]

Trachylina

An order of the class Hydrozoa of the phylum Coelenterata. The animals of this order are jellyfish of moderate size They differ from other hydrozoan jellyfish in having balancing organs which develop partly from the digestive epithelium, and in having only a small polyp stage or none at all. Many authorities now recognize three distinct orders of trachylines, and in this case the older term Trachylina is abandoned

Limnomedusae have a small polyp stage which produces other polyps and medusae by budding

tomatic recorder registered the presence or absence of the bats from their nests at night. In a similar manner the migration of cockroaches in New Orleans sewers has been studied, as well as the behavior of various insects and small animals.

The operation becomes more complex when a large number of biological particles are labeled, as, for example, in the tagging of red blood cells or bacteria. When the labeled substance is injected into an animal, it is impossible to follow the individual labeled particles, but their average movement can be tracked by observations of the radiation. In this way it is possible to measure the average lifetime of a red blood cell or the diffusion rate of bacteria

Finally, a radioisotope may be used to tag its own element. Phosphorus-32 can be introduced into the soil where a plant is growing, and the amount of phosphorus absorbed and its distribution throughout the plant studied.

In one of the very earliest tracer studies, G von Hevesy used radioactive lead to trace the pathways of lead in plants. It was essential in this study, as in any other tracer experiments of this type, that the chemical behavior of the radioisotope be identical to that of the stable isotope in order for it to circulate normally within the element being studied. For most isotopes this is essentially true. For very low-atomic-numbered elements, however, the mass difference between the isotopes may introduce a significant effect. This is particularly true in the case of hydrogen isotopes, in which the relative mass difference between radioactive and stable isotopes is quite large. Thus in many chemical and biological systems, deuterium and hydrogen will exhibit different effects. This is clearly illustrated by the fact that many biological species cannot exist when their H_2O is replaced by D_2O. Tritium, with a mass three times that of ordinary hydrogen, exhibits an even greater isotopic effect Experiments are often made first using deuterium, then tritium, as a label for hydrogen, and the different effects of the two experiments are noted. In the case of most elements, these isotope effects do not seriously limit the experiment.

In most biological tracer experiments, the radioisotope is introduced into the system and its radiation subsequently measured with Geiger-Müller counters or scintillation detectors (*see* GEIGER-MÜLLER COUNTER; SCINTILLATION COUNTER). Extremely soft (low intensity) radiations may be detected by the use of photographic film. In a typical experiment using this technique, part of a plant containing a radioactive isotope is sectioned and placed next to the film. The radiation from the plant darkens the film, and the pattern on the film shows the distribution of the radioisotopes in the plant. Such pictures are called autoradiographs (Fig. 1). They can be taken of cellular and even subcellular particles. Studies have been made of the reproduction of chromosomes using such techniques. Similar studies may be carried out in animals. *See* AUTORADIOGRAPHY; *see also* AGRICULTURAL SCIENCE (ANIMAL); PHOTOSYNTHESIS; PLANT, MINERAL NUTRITION OF.

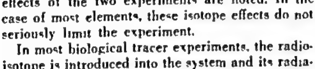


Fig. 1. Autoradiogram of metatarsus of 1-month-old calf 7 days following administration of calcium-45. Dense areas indicate highest concentration of the radioisotope. (From C L. Comar, Radioisotopes in Biology and Agriculture, McGraw-Hill, 1955)

Uses in medicine. Many medical tracer techniques have achieved the status of important and widely used diagnostic tests. For example, radioiodine has found wide use in determining the status of the human thyroid gland. In this test an isotope of iodine, usually iodine-131, is administered orally to a patient. After sufficient time has elapsed for the thyroid to assimilate the iodine, which is specific to that gland, a detecting device is used to measure the fraction of the iodine which has been absorbed by the thyroid gland. This fraction helps to determine whether the thyroid is overactive, normal, or underactive.

The water content of the body, as well as the total amount of certain electrolytes, may also be determined by tracers using a simple isotope dilution technique Deuterium and tritium are commonly used for water volume measurements, and sodium-24 and potassium-42 permit measurement of the sodium and potassium in the body which are freely exchangeable. *See* ISOTOPE DILUTION TECHNIQUES

There are various scanning techniques which enable one to locate a radioisotope within the body and consequently to delineate an organ or an abnormality within an organ. Scanners have been devised which can accurately fix the size of the thyroid gland and determine the presence of nodules or abnormalities in the shape of the gland. These scanners are also used to determine the location and extent of nodules of thyroid cancer which have been transplanted to other parts of the body

Other scanning techniques have been highly successful in locating tumors within the brain. The localization of brain tumors has been particularly successful because of a poorly understood process whereby normal brain tissue excludes a wide variety of isotopes and chemical compounds, or admits them slowly, while the tumor tissue takes them up to a much greater degree. The resultant difference between the isotopic concentration in the diseased tissue and in the healthy brain tissue permits the tumor to be spotted in many cases. Scanning techniques offer great promise in the detection of tumors and abnormalities in other organs (Fig 2).

Uses in industry. Radioactive tracers are becoming increasingly useful in industrial applications such as the testing of wear and corrosion in mechanical components. Piston ring wear within an engine can be determined accurately by transform-

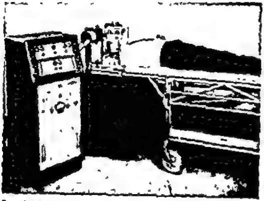


Fig. 2 Brain tumor scanner. (Baird-Atomic Co.)

ing some of the carbon within the ring into carbon-14. This is done by placing the ring in an intense neutron flux. The piston ring is then installed in the engine and, after a period of use, samples of the engine oil are removed and tested for carbon-14 content. In this way amounts of wear much smaller than could be found by weight measurements and other techniques can be determined.

On a larger scale, the entire operation of an industrial plant can be tested and controlled by tracer techniques. In one industrial survey sulfur-35 ...

... plant and its presence in the iron were studied by taking samples and measuring the iron's sulfur-35 content. It is entirely possible that similar techniques could be used in the routine operation and monitoring of many industrial processes. Such applications have been restricted in the past chiefly because of the difficulty of hazard control, but with adequate safeguards and more sensitive detecting devices, such procedures may soon become commonplace. [G.L.B.]

Bibliography: *See* RADIOACTIVITY (APPLICATIONS).

Tracheophyta

Plants having vascular systems consisting of specialized water and food-conducting tissues. Tracheophyta, a phylum of the subkingdom Embryophyta, contains over five-sevenths of all known plants or about 260,000 species, distributed throughout the world. This phylum includes club mosses, horsetails, ferns, conifers, and flowering plants. Despite variation in body form from tiny herbs to gigantic trees, all members have certain characteristics in common. They are fundamentally land plants although some, such as water ferns and water lilies, have gone back to an aquatic habitat. Each plant has xylem (water-conducting tissue) and phloem (food-conducting tissue). All have tracheids or such phylogenetic derivatives as vessel elements and wood fibers.

Reproduction. An alternation of generations is thought to occur in all members. The sporophyte (spore producing generation) is large, usually showy, and at maturity independent (can make its own food), whereas the alternating generation, the gametophyte (sex-cell-producing), is smaller, less conspicuous, and dependent.

Origin. The tracheophytes are believed to have evolved from Chlorophyta (green algae) which invaded the land from the sea (*see* CHLOROPHYTA). It has been suggested that at one stage in the evolutionary process, thalluslike species (structurally simple and frequently prostrate plants) differentiated forming roots, stems, and leaves.

Economic importance. Tracheophyta are of great economic importance. Products obtained from these plants include grain and forage, lumber, cork, textile fibers, cordage, oils, gums, resins, rubber, spices, perfumes, drugs, beverages, dyes, tanning materials, waxes, and paper.

The phylum Tracheophyta is composed of four subphyla: *Psilopsida*, *Lycopsida*, *Sphenopsida*, and *Pteropsida*. *See* LYCOPSIDA; PSILOPSIDA; PTEROPSIDA; SPHENOPSIDA; *see also* PLANT KINGDOM. [P.D.S.]

Bibliography: *See* EMBRYOPHYTA.

Trachoma

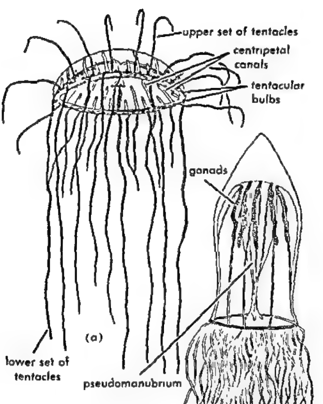
An infectious disease of the conjunctiva and cornea of eye caused by a virus of large size which resembles the psittacosis virus and the lymphogranuloma venereum virus. This serious granular conjunctivitis is of wide distribution. Cytoplasmic inclusions in the conjunctival cells consist of masses of particles resembling the elementary bodies of the psittacosis-lymphogranuloma venereum group. They stain reddish purple with Giemsa and blue with Castaneda's stain. The particles are held together by inclusion material which stains brown with iodine because of its glycogen content.

The disease is transmitted from eye to eye by fingers or fomites, possibly flies. It is usually acquired early in life by the child from the mother. There is inflammatory swelling of the upper eyelids and granulation of the tarsal conjunctiva may spread to the cornea. In the chronic stage, scarring causes deformity of the eyelids and vascularization or pannus of the cornea. Early oral treatment with sulfonamides or topically administered tetracyclines in ointments have reduced the incidence of the chronic stage in the Indians of the United States. *See* LYMPHOGRANULOMA-PSITTACOSIS GROUP. [K.F.M.E.]

Trachylina

An order of the class Hydrozoa of the phylum Coelenterata. The animals of this order are jellyfish of moderate size. They differ from other hydrozoan jellyfish in having balancing organs which develop partly from the digestive epithelium, and in having only a small polyp stage or none at all. Many authorities now recognize three distinct orders of trachylines, and in this case the older term Trachylina is abandoned.

Limnomedusae have a small polyp stage which produces other polyps and medusae by budding



Trachymedusae. (a) *Olindias*, from life, Bermuda. (b) *Aglantha*, from life, Puget Sound. (From L. H. Hyman, *The Invertebrates*, vol. 1, McGraw-Hill, 1940)

The tentacles are hollow. The best known Limnomedusae are *Gonionemus* and *Craspedacusta*. *Gonionemus* is widely used in zoology classes as a representative jellyfish, and is described in most textbooks. It is often abundant in salt-water ponds and bays. *Craspedacusta* lives in fresh water and occurs in nearly all parts of the world.

Trachymedusae and Narcomedusae are jellyfish of the open seas. Their tentacles, unlike those of Limnomedusae, have a solid core consisting of a single row of endodermal cells, and there is no polyp stage

Narcomedusae differ from both Limnomedusae and Trachymedusae in having broad and often

[S.CR.]

Bibliography: F. S. Russell, *The Medusae of the British Isles*, 1954.

Trachyte

A light-colored, aphanitic (very finely crystalline) rock of volcanic origin, composed largely of alkali feldspar with minor amounts of dark-colored (mafic) minerals (biotite, hornblende, or pyroxene). If sodic plagioclase (oligoclase or andesine) exceeds the quantity of alkali feldspar, the rock is called latite. Trachyte and latite are chemically equivalent to syenite and monzonite respectively.

Texture. The extremely fine-grained texture and more or less glassy material are due to rapid cooling and solidification of the lava. Large crystals *(phenocrysts) are commonly sprinkled liberally* through the dense rock giving it a porphyritic texture. These may be well-formed and 1–2 in. across They appear as glassy crystals of sanidine, and in addition small mafic phenocrysts may be present. In latite the phenocrysts are largely plagioclase. As the quantity of glass increases, these porphyric rocks pass into vitrophyre; and as the abundance of phenocrysts increases, these rocks pass into trachyte porphyry. *See* PORPHYRY; VOLCANIC GLASS.

The detailed features of trachyte are best studied microscopically. Sanidine and orthoclase are dominant over oligoclase in normal (potash) trachyte. In alkali (soda) trachyte, both alkali feldspar and mafics are soda-rich.

Composition. Brown biotite mica is the common mafic. It occurs as flakes which may be more or less resorbed by the liquid in the late stages of solidification so that only patches of dusty iron oxide remain Normal trachyte commonly carries somewhat corroded and resorbed hornblende or diopside. Alkali trachyte usually contains soda-rich amphibole (riebeckite, arfvedsonite, and barkevikite) or pyroxenes (aegirine-augite or aegirite) Zoned crystals with diopsidic cores and progressively more soda-rich margins are common

Either free silica (quartz, tridymite, or cristobalite) or feldspathoids (leucite, nepheline, or sodalite) may be present in small amounts With increase in free silica, the rock passes into rhyolite; and with increase in feldspathoids, it passes into phonolite Accessory minerals as tiny grains and crystals are magnetite, ilmenite, apatite, zircon, and sphene.

Structure. Streaked, banded, and fluidal structures due to flowage of the solidifying lava are commonly visible in many trachytes and may be detected by a parallel arrangement of tabular feldspar phenocrysts. A distinctive microscopic feature is trachytic texture in which the tiny, lath-shaped sanidine crystals of the rock matrix are in parallel arrangement and closely packed This rather uniform pattern is locally interrupted where the laths more or less deviate or wrap around the phenocrysts Orthophyric texture is common where tiny feldspar crystals show a stumpy or square outline

Occurrence and origin. Trachyte is not an abundant rock, but it is widespread It occurs as flows, tuffs, or small intrusives (dikes and sills) It may be associated with alkali rhyolite latite or phonolite

Trachyte is commonly considered to have been derived from a basaltic magma by differentiation a process involving removal in large quantities of early formed crystals rich in iron, magnesium and calcium A factor of importance in the formation of some trachyte is contamination of the original magma by incorporation of foreign rock material

The chemical transformation of andesite to trachyte may have occurred (in the solid state) where calcium was removed and sodium added metasomatically. This may explain the origin of some keratophyres (a variety of soda rich trachyte). *See* IGNEOUS ROCKS; MAGMA; PETROGRAPHIC PROVINCE, SPILITE. [C.A.CA.]

Tractor

A wheeled, self-propelled vehicle for hauling other vehicles or equipment and for operating the towed implements; also, a crawler which runs on an endless, self laid track and performs similar functions. *See* AUTOMOTIVE VEHICLE.

Farm tractor. A farm tractor is a multipurpose power unit. It has a drawbar for drawing tillage tools and a power take-off device for driving implements or operating a belt pulley as shown in the illustration

The acreage to be worked, type of crops grown, and the terrain all impose their requirements on tractor design Accordingly models vary in such details as power generated, weight, ground clearance, turning radius, and facilities for operating equipment All models can, however, be grouped under [illegible]

hi, [illegible]

is what counts, therefore, tractors are rated by the horsepower they deliver at the drawbar and at the belt On small models, the drawbar and belt horsepower may run as low as 10, on large models the drawbar horsepower runs as high as 74, while belt horsepower reaches about 89.

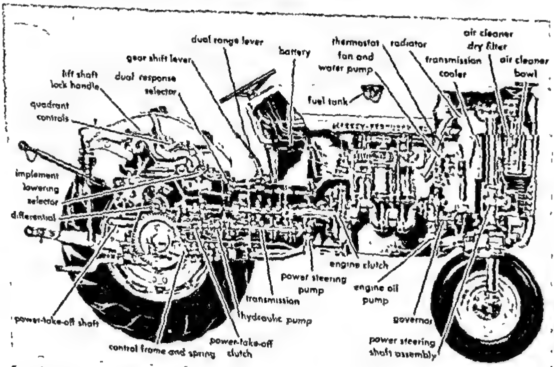
The major components are engine, clutch, and transmission. These components are intimately related and designed to work in conjunction with each other to accomplish specific work.

Unlike passenger-car engines, which are of the high speed type, tractor engines are relatively low speed; their maximum horsepower is generated at crankshaft speeds in the neighborhood of 2000 rpm. These engines have 1, 2, 3, 4, or 6 cylinders and operate on gasoline, kerosene, L-P gas, or diesel fuel. They are of the spark-ignition or diesel type, operating on the 4-stroke-cycle principle, and are cooled by water or air

Transmissions have 3, 4, 5, 6, 8, and 12 forward speeds and one or two reverse gears. Clutchless hydraulic transmissions are also used, making it possible to shift gears while in motion. Vehicle speeds are low, ranging from slightly more than 1 mph to about 14 mph in highest gear.

Interposed between transmission and engine is a single- or multiple-disk clutch of the general automotive type, manually or hydraulically controlled.

Power is taken off the transmission by shaft for operating equipment The power take-off may be manually controlled or hydraulically engaged and disengaged independently of vehicle speed. It may be run at one or two set speeds, depending upon the nature of the work it is intended to do. With pulley attached to the power take-off, a similar choice of speeds is afforded for belt work



Four wheel tractor showing major components from engine to power take-off and drawbar arms with hydraulic mechanism (Massey-Ferguson, Inc)

Hydraulic systems are used for control of both rear- and front-mounted implements. The rear-mounted implement may be attached to the two arms of the drawbar and to a third installed arm to give a 3-point hitch. An engine-driven hydraulic pump and cylinder built into the tractor provide the power to raise or lower the arms and thus to lift the implement to carry it, or lower it to the ground for work, at the will of the driver. The hydraulic mechanism may be so designed as to transfer weight from the front wheels to the rear wheels of the tractor as the load demands to give better traction when drawing multiple plows. Some systems are designed to control the depth of plow penetration automatically; others will disengage the clutch automatically if the mounted implement strikes an obstruction. The hydraulic cylinder may be attached to the implement, or may be part of it.

With the exception of crawlers, tractors are steered by turning a wheel. Hydraulic power devices are also used as manual assists. *See* STEERING, POWER.

A tractor has no frame and no springs. The supporting structure is a housing, or housings, for the transmission, clutch, hydraulic mechanism, drive shaft, and differential to which are bolted the engine crankcase, or frame for carrying the engine, and the housing for the rear axle The cushioning of the tractor load depends on the tires

A typical front axle is pivoted at its center to accommodate for rough ground Made in sections, it can be extended to change the width between wheels for straddling crop rows. The rear tread width can also be adjusted by assembling the wheel disk and rim in different positions.

Braking is accomplished through drum or disk brakes on the rear wheels, or by brakes on the differential. Both types permit the brakes to be applied independently on either side to assist in making sharp turns.

Crawler-type tractors in the smaller models are also used for farm work. In lieu of rubber-tired wheels, the crawler propels itself on two endless tracks made of hinged steel plates. Each track envelops a front and rear wheel; each is carried on a roller track frame which supports the weight of the tractor. The track is made to revolve by engaging teeth on the periphery of the rear wheels. Dual clutches permit power to be applied independently to each track. Steering is accomplished by manipulating levers which control the clutches to send power to one track while keeping the other motionless. *See* AGRICULTURAL MACHINERY.

Industrial tractor. The basic design of an industrial tractor for hauling and for operating construction equipment departs little from that of a farm tractor, and differences in design of models are for purposes of fitting the vehicle to the work it is intended to do

Because high ground clearance is not needed for industrial work, the tractor is commonly built with a lower center of gravity and capable of traveling a few miles per hour faster than a farm tractor. If its use is confined to hauling, it may not be equipped with hydraulic power. If it is to be used for operating a scraper, backhoe, or front-end loader, its structure may be heavier and more rugged *See* BULK-HANDLING MACHINES. [P.H.S.]

Tractrix

A plane curve for which the length of any tangent between the curve and a fixed line is constant c. If the x axis is the fixed line, its differential equation is

$$(dy/dx)^2 = y^2/(c^2 - y^2)$$

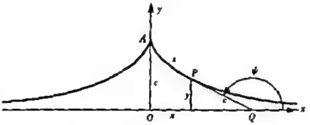
With ψ, the inclination of the tangent, as parameter, this yields the equations

$$x = c \log \tan \tfrac{1}{2}|\psi| + c \cos \psi \quad \text{and} \quad y = c \sin \psi$$

The tractrix has the x axis as asymptote and cusps at $(0, \pm c)$. The arc AP, measured from a cusp, is

$$s = -c \log \sin \psi = c \log \frac{c}{y}$$

A taut string PQ, moving in a horizontal plane so that Q describes a straight line, will pull a heavy particle at P along a tractrix The tractrix is an involute of a catenary.



Tractrix (upper half)

A tractrix revolved about its asymptote generates a pseudospherical surface of constant negative Gaussian curvature $-1/c^2$, its area is $4\pi c^2$, the same as a sphere of radius c *See* CATENARY

[L BR.]

Traffic control systems

Systems which act to control the movement of vehicles, such as airplanes, railroad trains, and automobiles. The system of traffic lights used in most urban areas to control the flow of automotive traffic is a common example of a traffic control system

Traffic control systems also are employed to insure safe and orderly arrival of airplanes at airports during periods of heavy traffic, to provide efficient, economical movement of railroad cars from one point to another, and to reduce congestion and provide a smooth flow of traffic along heavily traveled roads in metropolitan areas The nature of the traffic control system and its complexity are determined by the severity of the traffic problem and by the nature of the requirements placed upon the traffic control system.

Air-traffic control system. An air-traffic control system is required in most large airports, espe-

cially during conditions of poor visibility, to prevent collisions in the air or on the ground and to guide the airplanes in to a safe landing. Such systems are usually not automatic but require a human operator to instruct the airplane pilot. In many air-traffic control systems, the airplanes approaching the airport are "stacked up," that is, as an airplane arrives it is ordered to fly in a circular course over the airport at a specified altitude, which is higher than the assigned height of the last previous airplane. The airport traffic control manager then orders the airplanes to land one by one. As each airplane approaches for a landing, its altitude and direction are observed by radar (especially at night, or in inclement weather), and any necessary directions are radioed to the pilot during the landing. See AIR TRAFFIC CONTROL; NAVIGATION SYSTEMS, ELECTRONIC.

Railroad-traffic control system. The movement of railroad cars all over the country presents an interesting traffic control problem. Frequently, freight cars are loaded at one point and must then be dispatched to the points where they are unloaded and subsequently returned to the starting point, possibly after being reloaded at some intermediate point. To keep operating costs at a minimum, the routes of these freight cars are chosen in such a way that they travel the minimum required distance. Many modern railroads make extensive use of digital computers to choose the most efficient routing for their freight cars. See RAILROAD ENGINEERING.

Automobile-traffic control systems. Automobile traffic control systems range in complexity from the simple posting of speed limits, which is adequate for most lightly traveled suburban roads, to systems which employ computers to control the lengths of time that the red and green lights are on. The traffic control systems employed in Denver, Colorado and Baltimore, Maryland continuously measure the flow of traffic into and out of the city and employ these data to control the city's traffic lights. [J.C.TR.]

Bibliography: H. H. Goode and R. E. Machol, *System Engineering*, 1957.

Traffic engineering

Traffic engineering determines the required capacity and layout of highway and street facilities that can safely and economically serve vehicular movements between

sec

fu

pressways, the latter are also known as freeways. Secondary streets are sometimes described as collectors. Roads in rural areas are usually designated as primary, secondary, or tertiary, depending upon their importance.

In planning improvement in traffic arteries, consideration must be given to many factors, including traffic volumes, desired speeds of travel, vehicular dimensions, driving habits, vehicular performance, type of terrain traversed, and especially expected future changes in all of these factors.

Study of traffic movements. The traffic engineer must know both the pattern and volume of vehicular movements.

Traffic recorders. Mechanical counters or recorders are employed to determine traffic movements on an existing route. Four general types of counters are the air-impulse counter, magnetic detector, photoelectric counter, and radar detector. These devices are used to record hourly variations and total daily volumes of traffic at a given point.

Weigh stations are employed primarily to check against overloaded commercial vehicles. Operation of the stations also provides data on the average weight of the vehicles checked.

Origin-and destination studies. Information obtained by traffic counters and weigh stations is utilized in planning improvement, such as the widening or strengthening of an existing facility. When planning improvement of a whole network of roads or the construction of an entirely new system, additional engineering tools and procedures may be employed to evaluate traffic requirements. This more comprehensive investigation generally includes an origin-and-destination study. In this investigation house-to-house interviews with drivers may be conducted. To reduce this work and its cost only the residents of a previously selected percentage of the homes may be interviewed.

An origin-and-destination study also may be made by the cordon method. With this procedure on a selected day and for a specific period of time, all drivers are stopped and interviewed at the fringes of the area under investigation. Detailed data relating to time of trip, purpose of trip, route traveled, means of parking at destination, and desired improvements are secured.

Design volume. Traffic elements evaluated in planning important highway projects include: (1) current average daily traffic using the present facility, or that volume which would use a facility if it existed; (2) average daily traffic expected for a specified future year, which is generally 20 years or longer in the future; (3) anticipated directional distribution of the traffic at predicted peak hours; (4) expected ratio of trucks and other commercial vehicles to passenger cars; and (5) a design hour volume. Because the movement of traffic is not uniform throughout the day and night, street and highway improvements must be designed for peak flows, and the interval of time generally used for the design period is 1 hour. Experience has proven that if a system has adequate capacity to handle maximum flows for one hour the system will generally have sufficient capacity for longer periods.

Capacity calculations. With uninterrupted flow of traffic

be used. Next, those factors that may reduce the 1200 figure are evaluated, and this rate is scaled downward. By using the adjusted figure and know-

Hydraulic systems are used for control of both rear- and front-mounted implements. The rear-mounted implement may be attached to the two arms of the drawbar and to a third installed arm to give a 3-point hitch. An engine-driven hydraulic pump and cylinder built into the tractor provide the power to raise or lower the arms and thus to lift the implement to carry it, or lower it to the ground for work, at the will of the driver. The hydraulic mechanism may be so designed as to transfer weight from the front wheels to the rear wheels of the tractor as the load demands to give better traction when drawing multiple plows. Some systems are designed to control the depth of plow penetration automatically; others will disengage the clutch automatically if the mounted implement strikes an obstruction. The hydraulic cylinder may be attached to the implement, or may be part of it.

With the exception of crawlers, tractors are steered by turning a wheel. Hydraulic power devices are also used as manual assists. *See* STEERING, POWER.

A tractor has no frame and no springs. The supporting structure is a housing, or housings, for the transmission, clutch, hydraulic mechanism, drive shaft, and differential to which are bolted the engine crankcase, or frame for carrying the engine, and the housing for the rear axle. The cushioning of the tractor load depends on the tires.

A typical front axle is pivoted at its center to accommodate for rough ground. Made in sections, it can be extended to change the width between wheels for straddling crop rows. The rear tread width can also be adjusted by assembling the wheel disk and rim in different positions.

Braking is accomplished through drum or disk brakes on the rear wheels, or by brakes on the differential. Both types permit the brakes to be applied independently on either side to assist in making sharp turns.

Crawler-type tractors in the smaller models are also used for farm work In lieu of rubber-tired wheels, the crawler propels itself on two endless tracks made of hinged steel plates. Each track envelops a front and rear wheel; each is carried on a roller track frame which supports the weight of the tractor. The track is made to revolve by engaging teeth on the periphery of the rear wheels. Dual clutches permit power to be applied independently to each track. Steering is accomplished by manipulating levers which control the clutches to send power to one track while keeping the other motionless. *See* AGRICULTURAL MACHINERY.

Industrial tractor. The basic design of an industrial tractor for hauling and for operating construction equipment departs little from that of a farm tractor, and differences in design of models are for purposes of fitting the vehicle to the work it is intended to do.

Because high ground clearance is not needed for industrial work, the tractor is commonly built with a lower center of gravity and capable of traveling a few miles per hour faster than a farm tractor. If its use is confined to hauling, it may not be equipped with hydraulic power. If it is to be used for operating a scraper, backhoe, or front-end loader, its structure may be heavier and more rugged. *See* BULK-HANDLING MACHINES. [P.H S.]

Tractrix

A plane curve for which the length of any tangent between the curve and a fixed line is constant c. If the x axis is the fixed line, its differential equation is

$$(dy/dx)^2 = y^2/(c^2 - y^2)$$

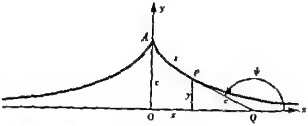
With ψ, the inclination of the tangent, as parameter, this yields the equations

$$x = c \log \tan \tfrac{1}{2}|\psi| + c \cos \psi \quad \text{and} \quad y = c \sin \psi$$

The tractrix has the x axis as asymptote and cusps at $(0,\pm c)$. The arc AP, measured from a cusp, is

$$s = -c \log \sin \psi = c \log \frac{c}{y}$$

A taut string PQ, moving in a horizontal plane so that Q describes a straight line, will pull a heavy



Tractrix (upper half)

particle at P along a tractrix. The tractrix is an involute of a catenary

A tractrix revolved about its asymptote generates a pseudospherical surface of constant negative Gaussian curvature $-1/c^2$; its area is $4\pi c^2$, the same as a sphere of radius c. *See* CATENARY.

[L.BR]

Traffic control systems

Systems which act to control the movement of vehicles, such as airplanes, railroad trains, and automobiles The system of traffic lights used in most urban areas to control the flow of automotive traffic is a common example of a traffic control system

Traffic control systems also are employed to insure safe and orderly arrival of airplanes at airports during periods of heavy traffic, to provide efficient, economical movement of railroad cars from one point to another, and to reduce congestion and provide a smooth flow of traffic along heavily traveled roads in metropolitan areas. The nature of the traffic control system and its complexity are determined by the severity of the traffic problem and by the nature of the requirements placed upon the traffic control system.

Air-traffic control system. An air-traffic control system is required in most large airports, espe-

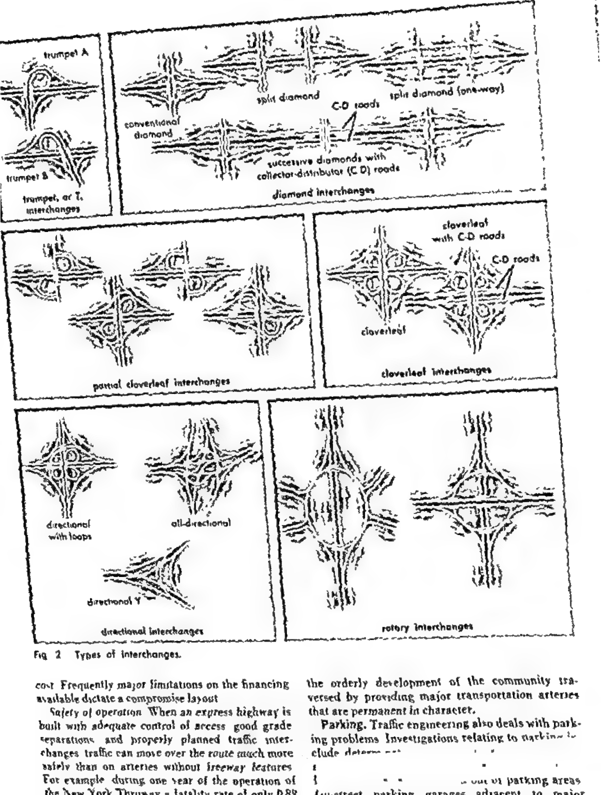


Fig 2 Types of interchanges.

cost Frequently major limitations on the financing available dictate a compromise layout

Safety of operation When an express highway is built with adequate control of access good grade separations and properly planned traffic interchanges traffic can move over the route much more safely than on arteries without freeway features For example during one year of the operation of the New York Thruway a fatality rate of only 0.88 deaths for each 100,000,000 miles of travel was established This rate was far lower than that on nearby roads not designed as freeways.

In addition to providing greater safety and more economical vehicle operation, control of access preserves the capacity of the road from future encroachments, thus safeguarding the investment in the expressway. Controlled access also assists in the orderly development of the community traversed by providing major transportation arteries that are permanent in character.

Parking. Traffic engineering also deals with parking problems Investigations relating to parking include deterr[illegible]

[illegible] out of parking areas Off-street parking garages adjacent to major streets may involve access connections requiring careful design In the central section of a city, flow to and from the off street parking areas may be so highly concentrated in short morning and evening periods that feasible access connections between principal streets and parking structures may become major problems. *See* TRANSPORTATION ENGINEERING

[A N C.]

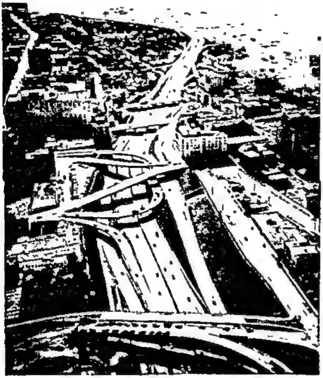


Fig 1. One of the most difficult highway engineering problems is the planning, financing, and construction of adequate expressways in large metropolitan areas. This Los Angeles, Calif., freeway cost several million dollars a mile.

ing the total design hour volume, the desired number of lanes is determined.

Capacity of a street or traffic lane at safe speeds is dependent upon many factors. These include available sight distances, curvature, maximum grades, width and character of roadway, shoulder construction, frequency of intersections, spacing of bus stops, clearances at bridges, location and frequency of loading and unloading zones, type of traffic control devices at intersections, whether or not left-turning movements are permitted at intersections, ratio of trucks and other commercial vehicles to passenger car volume, adequacy of directional signs, parking movements, type of lighting facilities employed, weather conditions, and whether or not traffic traveling in opposite directions is separated by a median or island.

Capacity is particularly influenced by the frequency of intersections of other arteries at grade. If only small volumes of traffic are carried on all streets, stop signs may not be needed at intersections. As traffic increases, stop signs must be provided on minor streets, and with still greater volumes power-operated control devices or signals must be installed at intersections. These signals are of two types: (1) fixed time, which produce a consistent and regularly repeated sequence of signal indications; and (2) traffic-actuated control devices in which the intervals of travel are varied in accordance with demands of traffic on the different streets as registered by the actuation of detectors.

Safe speeds. The frequency of side streets and the interference due to parking operations generally limit safe vehicular speeds on city routes to about 30 mph. In contrast, speeds up to 70 mph are permissible on express highways. These higher speeds are possible because grades are gradual, wide traffic lanes and good alignment are provided, sight distances are adequate, access to the highway is controlled, grade separations are used where other routes are crossed, and a traffic interchange is employed where the expressway intersects another important artery.

Design standards. Where a high-speed major highway across flat open country is being planned, traffic engineering indicates that major benefits will follow if uphill grades of the main roadways are limited to 3%, each traffic lane is made at least 12 ft wide, maximum horizontal curvature is restricted to 5° to provide long sight distances, and vertical curves are at least 600 ft in length to assist in securing adequate sight distances. In rolling and mountainous country economy may dictate that these standards be reduced.

Access control. Control of access limits the number of points of entrance and departure from the expressway, thereby increasing its capacity. Access is by entrance ramps incorporating an acceleration lane that permits the entering vehicle to reach the speed of those already upon the expressway before moving onto the main roadway. Exit from the expressway is by an off-ramp incorporating a deceleration lane. Normally a distance of several hundred feet between ramps is specified to provide adequate distances for the streams of traffic to intermingle.

Grade separations are bridges or structures used to separate vertically two intersecting roadways, thus permitting traffic on the one road to cross traffic on the other road without interference. They are also called overpasses and underpasses.

Traffic interchanges. A traffic interchange is a system of interconnecting roadways in conjunction with a grade separation or separations providing for the interchange of traffic between two or more roadways or highways on different levels (Fig. 1). Major types of traffic interchanges include directional, cloverleaf, diamond, rotary, trumpet, and variations of these basic types (Fig. 2). An advantage of the simple directional Y interchange is that high speeds of travel may be permitted on all roadways, while with more complicated directional layouts and with cloverleaf interchanges reduced speeds on the ramps are usually necessary. Left-turning movements and need for control signals are disadvantages of the diamond interchange, but this type has the advantage of occupying a minimum area. The rotary type lends itself for use when several streets intersect, but it requires considerable weaving of vehicles, which slows traffic movements and reduces capacity of the interchange.

Factors influencing the type of interchange selected include the expected volumes of traffic to be accommodated, the number of roads or streets involved, the distribution of the total traffic on the various roads, the land available for the interchange, topography, and the estimated construction

tor and glandular response as well as psychological features such as tension, attention, and perception. Subtle chemical differences among tranquilizers are revealing the paths by which brain systems are influenced, and subtle differences in the organization of well-known behavioral states such as fear, attention, excitement, and sleep.

Use in specific diseases. The tranquilizers that act on systems which influence the regulation both of mental alertness and tension, and their body correlates, are used in treating psychosomatic disorders, schizophrenia, and impulse disorders of senility, and childhood. Every new therapy in psychiatry makes it clear that deterioration and dilapidation are not inevitable consequences of mental disease, that there is a considerable recovery rate, and that the instigation and organization of a single disease or behavior is analyzable into a number of interacting factors over which knowledge and control can emerge. As chemical therapies restore behavioral equilibrium, behavioral science is challenged to find social and psychological values and treatment which facilitate and maintain personal health, rehabilitation, and development. [D.X.F.]

Bibliography: J. O. Cole and R. W. Gerard (eds.), *Psychopharmacology: Problems in Evaluation*, Natl. Acad. Sci.–Natl. Research Council, Publ. 583, 1959.

Transamination

The transfer of an amino group from one amino acid to another, without the formation of free ammonia or ammonium ion as an intermediate. The transamination reactions are essential components of many biosynthetic and degradative pathways. They also serve to ensure a balanced distribution of amino nitrogen in the living organism without its loss in the form of ammonia. It may occur nonenzymatically under special conditions, but in physiological systems it is catalyzed by one of a group of enzymes called transaminases. The metabolically important transaminations between glutamic acid and pyruvic acid and glutamic acid and oxaloacetic acid serve as examples:

$$\underset{\text{Glutamic acid}}{HOOC{-}CH_2{-}CH_2{-}\underset{\substack{|\\ NH_2}}{CH}{-}COOH} + \underset{\text{Pyruvic acid}}{CH_3{-}\underset{\substack{\|\\ O}}{C}{-}COOH}$$

$$\Updownarrow$$

$$\underset{\alpha\text{-Ketoglutaric acid}}{HOOC{-}CH_2{-}CH_2{-}\underset{\substack{\|\\ O}}{C}{-}COOH} + \underset{\text{Alanine}}{CH_3{-}\underset{\substack{|\\ NH_2}}{CH}{-}COOH}$$

$$\text{Glutamic acid} + \underset{\text{Oxaloacetic acid}}{HOOC{-}CH_2{-}\underset{\substack{\|\\ O}}{C}{-}COOH}$$

$$\Updownarrow$$

$$\alpha\text{-Ketoglutaric acid} + \underset{\text{Aspartic acid}}{HOOC{-}CH_2{-}\underset{\substack{|\\ NH_2}}{CH}{-}COOH}$$

Every known enzymatic transamination requires pyridoxal phosphate or pyridoxamine phosphate as coenzyme. The protein portion of an enzyme, the apoenzyme, catalyzes the transfer of the amino group from the donor compound to pyridoxal phosphate, forming pyridoxamine phosphate, and then transfers the amino group from the latter to the substrate acceptor, regenerating pyridoxal phosphate. Thus, the above example is more correctly written:

$$\begin{array}{c} \text{Glutamic acid} \\ \\ \alpha\text{-Ketoglutaric acid} \end{array} \Bigg) \Bigg(\begin{array}{c} \text{Pyridoxal phosphate} \\ \\ \text{Pyridoxamine phosphate} \end{array} \Bigg) \Bigg(\begin{array}{c} \text{Alanine} \\ \\ \text{Pyruvic acid} \end{array}$$

All naturally occurring α-amino acids can transaminate with α-ketoglutarate. A trace of glutamate or α-ketoglutarate will thus catalyze transamination between any two amino acids, provided the necessary transaminases are present, as shown in the equation below.

A few transaminases have been highly purified and found to be relatively specific, but in most cases it is not known whether the observed transamination is a direct transfer or the sum of two reactions, as in the case illustrated above. In other instances a transaminase has been found to catalyze amino transfer between any two members of a group of compounds. For example, the so-called transaminase A of *Escherichia coli* transfers the amino group from any one of the following to the keto acid corresponding to any one of the others: glutamic acid, aspartic acid, tryptophan, tyrosine, phenylalanine, and methionine.

In addition to α-amino acids, certain ω-amino acids (such as ornithine, glycine, and γ-aminobutyric acid) as well as purines and pyrimidines may transaminate with α-ketoglutaric acid. See ALANINE; AMINO ACIDS; ASPARTIC ACID; ENZYME; GLUTAMIC ACID; METHIONINE; PHENYLALANINE; TRYPTOPHAN; TYROSINE. [E.A.AD.]

Bibliography: A. Meister, *Biochemistry of the Amino Acids*, 1957.

$$\text{Amino acid(1)} + \alpha\text{-Ketoglutaric acid} \underset{}{\overset{\text{Transaminase(1)}}{\rightleftharpoons}} \text{Keto acid(1)} + \text{Glutamic acid}$$

$$\underline{\text{Glutamic acid} + \text{Keto acid(2)}} \overset{\text{Transaminase(2)}}{\rightleftharpoons} \underline{\alpha\text{-Ketoglutaric acid} + \text{Amino acid(2)}}$$

$$\text{Sum}\quad \text{Amino acid(1)} + \text{Keto acid(2)} \rightleftharpoons \text{Keto acid(1)} + \text{Amino acid(2)}$$

Trajectory

The curve described by a body moving through space, as of a *meteor through the atmosphere, a* planet around the sun, a projectile fired from a gun, or a rocket in flight. In general, the trajectory of a body in a gravitational field is a conic section—ellipse, hyperbola, or parabola—depending on the energy of motion. The trajectory of a shell or rocket fired from the ground is a portion of an ellipse with the earth's center as one focus; however, if the altitude reached is not great, the effect of gravity is essentially constant, and the parabola is a good approximation. *See* BALLISTICS, EXTERIOR. [J.P.H.]

Tranquilizer

A class of psychopharmacologic drugs that tend to calm overexcited patients, and are unique in inducing drowsiness without impairing ready arousal, *and in restraining hyperactivity without inducing* coma or arrest of respiratory muscles. The effects sometimes encountered in other drug and organic therapies are minimal in tranquilizer therapy. For example, direct impairment of initiative and induction of impulsivity, amnesia, confusion, or thought deficit do not occur. Furthermore, the most potent tranquilizers do not induce addiction, withdrawal convulsions, or toxic psychoses. Their potency and relative safety have stimulated therapeutic activity in the mental hospital and have initiated research into the neurochemical basis for the array of unique effects. Synonyms proposed for tranquilizers are ataractic (peace of mind), neuroleptic, and psychotropic, indicating notions about mechanism of action or effect. *See* PSYCHOPHARMACOLOGIC DRUGS, PSYCHOSIS.

Mechanism. Neurochemical systems through which tranquilizers act are either unidentified or not linked specifically with the mental disease which the drugs ameliorate, but cannot be said specifically to cure. They appear to moderate the ease with which excitement and associated mental *anguish are evoked. Most clearly, they inhibit, but* do not abolish, motor expressions of mania, excessive irritability, and tension. It is less clear whether these reduced motor reactions also involve a preceding, simultaneous, or secondary reduction of painful emotion and attention to it.

Effect on patient. *The treated psychotic patient* who shows restored control and organization may be calm but still deluded. Psychotherapy and insight are not necessarily facilitated. Behavioral improvement occurs and the attendant diminution in bizarre performance leads to improved relations with ward personnel and *to better attention better* received. This in turn indirectly diminishes damaging excesses or self-neglect, impulsiveness, and isolation. With community and family acceptance, and with judiciously maintained drug and psychotherapy, the patient may adjust successfully, and readmission and first admission rates may *be reduced.* Tranquilizers do not eliminate the usefulness of older organic therapies but have replaced them for some patients. They have greatly enlarged the treat*ment possibilities available for any one patient. See* PSYCHOTHERAPY.

Tranquilizer drugs are effective with pathological intensities in behavior rather than with the tensions of everyday life. Depression and withdrawal are produced in susceptible persons, and undesired side effects can occur. According to the drug used these may include vasomotor shock, liver, blood, skin, and visceral damage. Not uncommonly a syndrome of compulsive restlessness, or akathisia, occurs and this is easily mistaken for mental disease. In large doses, muscle spasticity and tremor appear.

Types of tranquilizers. According to basic chemical structure, four categories of tranquilizers are delineated: phenothiazines, for example, chlorpromazine; rauwolfia derivatives, for example, reser*pine, propanediols, for example, meprobamate or* Miltown, diphenylmethanes, for example, benactyzine; and a miscellaneous group which includes ectylurea, glutethimide, and methylparafynol.

Rauwolfia root was used in India as early as 1000 B.C. and bromide, paraldehyde, and barbiturate sedatives were introduced between 1857 and 1903 (*see* BROMIDE; RAUWOLFIA). In 1952 chlorpromazine, and in 1953, reserpine, were adopted in psychiatric practice as tranquilizers. They had initially been developed in the laboratory for entirely different effects such as those on blood pressure, temperature, and allergic reaction. Phenothiazines prevent some peripheral actions of adrenaline, a hormone which is released into the blood stream during stress, but explanations for their central action currently lie in neurophysiologic studies. Reserpine induces the release from intracellular storage depots of a number of hormones such as serotonin, and noradrenalin which influence the tone and reactivity of smooth muscle, and perhaps the response of neurones. Such findings have led to studies of the metabolic and neural function of th*ese hormones and a search for their qualitative* and quantitative differences in health and in disease. *See* SEROTONIN.

The brain reflects graded states of excitement or sedation with a heightened or lowered level of general activity in a core of deep-lying areas in which *neural nets are discharging indirectly to the cortex* and rather directly to glands, blood vessels and muscles. Shifts in the levels of summated activity of this core, the reticular systems, are influenced by hormones and chemicals, by input from the body and environment, by intrinsic rhythms, and by other *brain connections. Different drugs act on these different* components to produce tranquilization. By regulating the readiness with which distant sensory and motor neurones receive or discharge an impulse, these reticular systems affect the degree and duration of arousal both of consciousness, and its *expression. They also may reflect the degree of mo-*

terephthalic acid. Monoglycerides of unsaturated fatty acids are prepared by alcoholysis of the appropriate oil with glycerol. The properties of lard may be improved by combining ester interchange among the natural mixture of glycerides with precipitation of the higher-melting material. In this manner the saturated glycerides are preferentially removed.

Interesterification reactions are generally reversible and accompanied by smaller heat effects than are other esterification reactions. The reactions are conducted in the liquid phase, usually in the presence of a catalyst. With catalysts, the common reaction temperature is about 100°C; without a catalyst, it is as a rule about 250°C. One major exception is the directed ester interchange in fats, which is conducted at temperatures around 50°C. The lower temperature permits crystallization of the saturated glycerides. A pressure of one atmosphere is normal, and the reaction is pushed to higher conversions by using an excess of the displacing acid or alcohol.

Reaction equilibria. The *equilibrium constants*, and thus the equilibrium conversion, can be predicted from the data for the formation of the ester reactants and products. Consider the alcoholysis reaction

$$\mathrm{RCOOR' + R''OH \rightleftharpoons RCOOR'' + R'OH} \qquad (1)$$

The reactions leading to the formation of the two esters from the appropriate acid and alcohol are

$$\mathrm{RCOOH + R'OH \rightleftharpoons RCOOR' + H_2O} \qquad (2)$$

$$\mathrm{RCOOH + R''OH \rightleftharpoons RCOOR'' + H_2O} \qquad (3)$$

The respective concentration equilibrium constants are

$$K_1 = \frac{[\mathrm{RCOOR''}][\mathrm{R'OH}]}{[\mathrm{RCOOR'}][\mathrm{R''OH}]} \qquad K_2 = \frac{[\mathrm{RCOOR'}][\mathrm{H_2O}]}{[\mathrm{RCOOH}][\mathrm{R'OH}]}$$

$$K_3 = \frac{[\mathrm{RCOOR''}][\mathrm{H_2O}]}{[\mathrm{RCOOH}][\mathrm{R''OH}]}$$

so that the equilibrium constant for the alcoholysis reaction may be obtained from the ratio of the equilibrium constants (K_3/K_2) for the two esterification reactions. If R″ is an alcohol with a large esterification equilibrium constant relative to R′ for the same acid, RCOOH, a large value will be obtained for K_1. The reaction will be essentially irreversible.

Reaction kinetics and catalysis. The rates of transesterification reactions respond to reaction conditions and molecular structure in a manner similar to esterification reactions in general. Catalysts useful for esterification also are good for transesterification. In the presence of an alkaline catalyst, the rates of alcoholysis are very high, being faster even than saponification. Sodium alkoxides are particularly active for alcoholysis and ester interchange and permit practical reaction rates to be obtained at lower temperatures than those of esterification. Because of their basic nature, these catalysts obviously would be unsuitable for other esterifications.

Reaction mechanism. Acid-catalyzed and base-catalyzed *transesterification reactions are explained by different mechanisms. Acid-catalyzed reactions follow the same course as other esterification* reactions. With base-catalyzed reactions, the basicity of the oxygen of the alcohol appears to be increased so that it attacks the *carbonyl carbon* directly. Electron-attracting substituent *groups in* either the acyl or the alkoxy part of the ester increase the reaction rate. Lower activation *energies are observed in base-catalyzed reactions. See* ACIDOLYSIS; ALCOHOLYSIS; ESTER; HYDROLYSIS.

[J. M.WO.]

Transference number

The *relative migration-velocities* of the anion and cation of an *electrolyte*. When an electric current is passed through a solution of an electrolyte, the various ions present carry different portions of the current. The proportions, which are related to the migration-velocities of the ions, are called transference numbers. (In British publications they are usually referred to as transport numbers.)

It can be shown that

$$\Lambda = F\alpha\,(U^+ + U^-) \qquad (1)$$

in which Λ is the *equivalent* conductance of an electrolyte yielding only two types of ions, F is the faraday, α is the degree of dissociation, and U^+ and U^- are the mobilities of the positive and negative ion constituents respectively. See ELECTROLYTIC CONDUCTANCE. The portion of the conductance $F\alpha U^+$ is due to the positive ion constituent, and $F\alpha U^-$ to the negative constituent. Thus the transference numbers, t^+ and t^-, are

$$t^+ = U^+/(U^+ + U^-) \quad \text{and} \quad t^- = U^-/(U^+ + U^-) \qquad (2)$$

In general the transference number on an ion, i, with the concentration C_i and the mobility U_i is given by the relation $t_i = C_iU_i/\Sigma C_jU_j$ in which the summation, Σ, is over all the j types of ion in the solution. This equation is applicable to mixtures of electrolytes.

The [illegible]
num[illegible]
liter[illegible]
and [illegible]
but [illegible] scientific interest is a third method, which depends upon measurements of the potentials of concentration cells with and without transference.

The Hittorf method. This can be illustrated by an example, shown in Fig. 1, of the determination of the transference numbers of silver nitrate. The two silver electrodes are placed in a solution of silver nitrate contained in the tube, and an electric current is passed in the direction indicated. The electrode reactions are $Ag = Ag^+ + e^-$ at the anode and $Ag^+ + e^- = Ag$ at the cathode (e^- represents the electron). For every faraday F of electricity passed, a full equivalent of silver ion constituent

Transducer

Any device or element which converts input energy into output energy of another form. An example is the microphone, which converts impinging sound energy into electric energy. In instrumentation the term is widely used, not only for intermediate elements of a system, but also for primary or sensing elements where identification of the input energy is not always easy or useful. In line with this common use of the term, the definition requires modification: a transducer is any device or element the output of which differs in kind from the input, and in which the output is in a known relation to the input so that measurement of the input is possible. *See* INSTRUMENTATION.

The most important class of transducer in instrumentation is the one in which the output is electrical in nature, because of the ease with which transmission and amplification of the output is accomplished. Common examples are electrical resistances, inductances, and capacitances, where a primary element by its deflection is made to control the electrical quantity and thus its output. Other examples are the piezoelectric crystal, which gives an electrical output when subjected to a force or pressure; photoelectric cells, which produce an electric voltage proportional to incident radiation or light; and a hot wire, used to measure air-speed fluctuations, where the conductivity of heat away from the wire is proportional to the air flow, resulting in a change in electrical resistance of the wire. The latter example illustrates the difficulty of applying a definition of transducer based upon conversion of energy.

Transducers with a mechanical output are largely restricted, except for the microphone, to primary sensing elements. These are not commonly called transducers. [W.C.B.]

Transducer, underwater

A device used for the generation or reception of underwater sounds. The word projector is applied to a generator of sound, while hydrophone refers to a receiver. Since the same device may serve both purposes, the word transducer is used as a general designation (*see* TRANSDUCER). Thus, a transducer, when a projector, converts electrical energy to motional mechanical energy, whereas a hydrophone converts mechanical energy to electrical energy. This conversion of energy is usually based on one of the following properties of certain materials: piezoelectricity, magnetostriction, and electrostriction. In each case, the application of an alternating electric or magnetic field of a given frequency causes a mechanical vibration of the transducer material at the same frequency. The transducer, being in contact with the water, communicates a similar motion to it, giving rise to a sound wave. With a hydrophone, the vibratory sound pressure causes a mechanical motion of the material, which in turn generates an electrical voltage. This voltage is amplified, then either read on a meter, recorded, or played through a loudspeaker.

Common materials used are quartz and Rochelle salt for piezoelectric transducers, nickel and permalloy for magnetostrictive ones, and barium titanate for the electrostrictive types. *See* ELECTROSTRICTION; MAGNETOSTRICTION; PIEZOELECTRICITY.

An important characteristic of a transducer is its frequency response, by which is meant the dependence of the output amplitude upon the frequency of the input signal. In a resonant transducer, this response is significant only for frequencies around a characteristic resonant frequency. By appropriate design (that is, loading, backing, etc.), a transducer may be made to have a nearly uniform response over a considerable frequency range. The desirable frequency response in a transducer depends upon its application. To receive or send signals of a nearly single frequency, as in an echo sounder, a narrow-band frequency response is employed in order to get greatest sensitivity and efficiency (*see* ECHO SOUNDER). Other applications may demand a wide-band frequency response, as in the reception of an explosive pulse.

Another important property of a transducer is its directivity, or directional response. This measures the relative response (or output) of the transducer for waves in different directions, the frequency being fixed. The directivity is determined by the geometrical shape of the transducer, and if composed of an array or mosaic of individual elements, may also be controlled by adjusting the electrical phasing of these elements. A common configuration for a transducer used in echo-ranging sonar is a plane circular radiator which produces a searchlight-type beam pattern. Such a transducer is then trained in the desired direction. A common type of receiving hydrophone is a line hydrophone which has the shape of a long circular cylinder. This has greatest response for sound striking parallel to the axis and is independent of the angle about the axis. *See* DIRECTIVITY; SONAR; UNDERWATER SOUND. [R.W.MO.]

Bibliography: H. F. Olson, *Acoustical Engineering*, 3d ed., 1957.

Transesterification

The group of reactions in which an ester reacts with another compound to form a different ester. The other compound may be an acid, an alcohol, or another ester. These reactions are also known as acidolysis, alcoholysis, and ester interchange, respectively. They are all special cases of esterification and have the characteristics of esterification reactions in general. *See* ESTERIFICATION.

A few of the more important applications of transesterification follow. Acrylic and methylacrylic esters of high-molecular-weight saturated and unsaturated alcohols are prepared by alcoholysis of methyl acrylate and methyl methacrylate; polyvinyl alcohol, by alcoholysis of polyvinyl acetate with methyl alcohol; and the prepolymer of polyethylene terephthalate, by alcoholysis of the methyl ester of terephthalic acid with ethylene glycol, because less effort is required to prepare dimethylterephthalate of the required [illegible], than

excess of the solution under examination (leading solution) so that a drop protrudes. Now if the upper disk slides over the lower disk the excess amounts of both solutions are sheared away, and a boundary, little disturbed by mixing or diffusion, results when the tubes are in place over each other as is shown in Fig 3b.

In addition there is the "autogenic" boundary, first used by E. C. Franklin and H. P. Cady. The solution of electrolyte whose transference numbers are desired is placed in a tube over the bottom of which is placed a disk which forms a soluble salt in combination with the anion of the solution. For instance, the solution may be potassium chloride and the metal cadmium. Now if current is passed with the metal as anode, cadmium chloride will form at the metal surface, and a boundary between the potassium and cadmium ion constituents will move up the tube. The method is more restricted in its application than that using sheared boundaries, but its simplicity recommends its use when possible.

Even if it is quite sharp when initially formed a boundary would become indistinct from diffusion and convection if it were not for an adjusting effect which operates to overcome the results of such mixing. An example is shown in Fig. 4 in which a lithium chloride indicator solution follows a potassium chloride solution. Since the lithium ion constituent has a lower mobility than that of the potassium ion and since the lithium chloride solution is the more dilute, the passage of current will cause a greater potential drop in the former than in the latter. This is shown diagrammatically in the figure where values of the electromotive force are plotted as ordinates and distances along the measuring tube as abscissae. Now, if some of the relatively fast moving potassium ions diffuse or are carried by convection into the lithium chloride region they

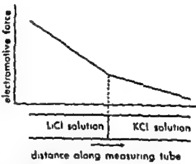


Fig. 4 Use of indicator solutions in moving boundary measurements.

will encounter a high potential gradient and will be rapidly sent forward to the boundary. Also, if lithium ions diffuse into the potassium chloride they will move more slowly than the potassium ions and will be overtaken by the boundary. Tests made using a highly colored indicator ion showed that a boundary, which had been allowed to become quite diffuse by stopping the current, rapidly regained its sharpness and, in addition, that the diffuse boundary moved at the same rate as a sharp one, while in the process of reforming. A result of this rectifying effect is to produce a relation between the concentrations of the leading and indicator solutions represented by

$$t_1/C_1 = t_i/C_i \tag{5}$$

in which t_1 and t_i are the transference numbers of the leading and indicator ions and C_1 and C_i are the corresponding concentrations. Under these conditions the ions in the leading and following solutions are moving at the same rate. However, if the initial concentrations in the mixed solutions at the boundary are too far from those represented by Eq. 5, adjustment to this condition may not take place.

In the differential moving boundary method, a boundary between two different concentrations, C_1 and C_2, of the same electrolyte, such as 0.2 *N* LiCl and 0.5 *N* LiCl, will move, on passage of electric current, if the transference numbers t_1 and t_2 are different. The relation involved is

$$\Delta t = \Delta v \, \Delta C \tag{6}$$

in which Δt is the change of the transference number in the concentration range ΔC and Δv is the volume swept through per faraday *F*. To obtain transference numbers, rather than differences in those numbers, it is necessary to start with known values, usually determined by the direct method. The differential method is useful for concentrated solutions where the direct method may encounter difficulties because of relatively large Joule heating effects. Since the rectifying effect is not operative in this case, the boundaries are not as easily visible as those found with the direct moving boundary method. L. G. Longsworth, who has obtained the only quantitative results with the differential method used "schlieren scanning," which depends

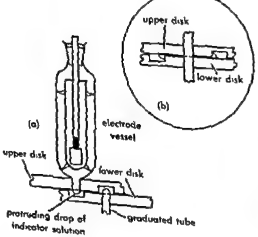


Fig. 3 Moving boundary transference number cell. (a) Method of filling. (b) Tubes in position for passage of current.

will appear around the anode and a like amount of that constituent will disappear from the region of the cathode. Between the electrodes, however, the current is only partially carried by silver ions, the conductance being shared in this portion of the cell by the negatively charged nitrate ions. Since the silver ions do not carry all the current, they do not move away from the region of the anode as fast as they form, and silver ions accumulate around that electrode. At the same time, the silver ion concentration is reduced around the cathode, since silver ions do not move into this region as fast as they are removed at the cathode. As seen in the figure, the tube is arbitrarily divided into three portions, of which the anode portion becomes more concentrated, the cathode portion more dilute, and the middle portion remains unchanged in composition, as the current passes. Since the silver and the nitrate ions carry all the current, the sum of the two transference numbers must be unity, that is, $t_{Ag^+} + t_{NO_3^-} = 1$. The changes around the anode, per faraday passed, may be summarized as follows:

Gained:

1 equiv. Ag^+ ion (formed from anode)
$t_{NO_3^-}$ equiv NO_3^- ion
(migrated from middle portion)

Lost:

$t_{Ag^+} = (1 - t_{NO_3^-})$ equiv. Ag^+ ion
(migrated to middle portion)

Net change:

$t_{NO_3^-}$ equiv. Ag^+ gain
$t_{NO_3^-}$ equiv NO_3^- gain

Thus the solution remains electrically neutral. A similar study of the changes at the cathode indicates that there is a loss, per faraday, of $t_{NO_3^-}$ equivalent of silver nitrate. To obtain the Hittorf transference number the changes in composition are referred to the solvent, usually water. Analysis of an electrode portion of solution yields the amount of solvent as well as of the solute, and the change in composition is referred to the amount of solute originally contained in this weight of solvent.

For the general case let N_O and N_F represent the original and final number of equivalents of an ion constituent associated with a given weight of solvent. Now if N_E is the number of equivalents of this ion added to this amount of solvent by the electrode reaction and tN_E the number lost by ionic migration (t being a transference number and N_E being also the number of faradays passed) then

$$N_F - N_O = N_E - tN_E$$

from which *the transference number is*

$$t = (N_E + N_O - N_F)/N_E \qquad (3)$$

With appropriate changes of sign this equation applies to both anode and cathode reactions.

Though the Hittorf method for obtaining transference numbers is simple and direct, it is difficult to obtain precise results with its use. The reason for this is that the method requires extremely accurate

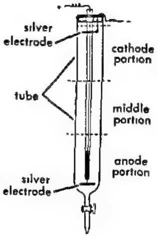
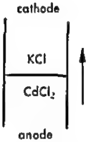


Fig 1. Hittorf transference number cell.

Fig. 2. Moving boundary between two solutions.



analytical procedures, since the results depend upon small differences between relatively large quantities.

The moving boundary method. This method has been developed to yield data of high accuracy. The phenomenon which makes the measurement possible may be illustrated by the following example. If a potassium chloride solution is placed in a tube above a cadmium chloride solution, as is shown in Fig. 2, and electric current is passed between electrodes placed with the cathode at top of the tube, the boundary will become sharp and will move up the tube. The motion of the boundary can be readily followed because the two solutions have different indices of refraction. The method is applicable to any ion constituent provided an indicator solution can be found. In the case just considered, it is the cadmium chloride solution. The mobility of the ion of the indicator must be lower than that of the leading ion, and its solution must be more dense for a rising boundary and less dense for a descending one. Thus a lithium chloride solution may be used to form a boundary for potassium chloride if the boundary descends. By measuring the volume swept through by the boundary during the passage of a given quantity of electricity, the transference number of the ion constituent ahead of the boundary may be obtained with the aid of the equation

$$t = vCF/sI \qquad (4)$$

in which v is the volume swept through by the boundary in s seconds while I amperes are passing, C is the concentration in equivalents in unit volume, and F is the faraday.

For the successful use of the method it is necessary to start with an adequately sharp boundary, *that is, the leading and indicator solutions must not* be too much mixed. The device for forming a "sheared" descending boundary, devised by D. A. MacInnes and T. B. Brighton, is shown in Fig. 3a. The electrode vessel is filled with the indicator solution and is closed with the stopper which carries the electrode in such a way that a drop of indicator solution hangs from the open end of the tube. The graduated tube is filled with a slight

can be found in which A is represented by a diagonal matrix. Of course the basis which diagonalizes the operator A is the set of eigenfunctions $u(\alpha)$ of A, and the diagonal values of A with this basis are its eigenvalues α. Moreover, in matrix equations, for example, the equation of motion of A

$$\frac{dA}{dt} = \frac{1}{i\hbar}(AH - HA) \tag{1a}$$

remain invariant under unitary transformation, as is easily verified.

Changes of basis in quantum theory often are termed canonical transformations, because they correspond to transformations which in classical mechanics leave invariant Hamilton's equations of motion. *See* CANONICAL TRANSFORMATIONS; HAMILTON'S EQUATIONS OF MOTION.

Since the Schrödinger equation conserves total probability, $\psi^\dagger(t)\psi(t) = \psi^\dagger(0)\psi(0)$, so that the evolution of the wave function in time is a continuous rotation in Hilbert space. In fact the Schrödinger representation can be interpreted as holding the basis fixed, whereas the Heisenberg representation keeps the state vector fixed in Hilbert space but rotates the basis $u_n(0)$ to $u_n(t)$.

Density matrix. To any normalized wave function ψ there corresponds a matrix $\rho_{mn} = \psi_m\psi_n^*$, known as the density matrix. The matrix ρ is Hermitian, $\rho_{nm}^* = \rho_{mn}$, and it transforms just as any other matrix on a change of basis, that is, $\rho' = S\rho S^\dagger$. Instead of Eq. (1a), however,

$$\frac{d\rho}{dt} = -\frac{1}{i\hbar}(\rho H - H\rho) \tag{1b}$$

The density matrix associated with ψ has the very special property that in the basis where ρ is diagonal all elements of ρ vanish, except a single diagonal element whose value is unity; obviously all projections ψ_m are zero in this basis, except a single projection whose square is unity. This formal result, which means simply that given any wave function ψ there is an operator A for which ψ is an eigenfunction, has profound implications. For example, suppose $\psi_+(\mathbf{r})$, $\psi_-(\mathbf{r})$ are the two components of the wave function of an electron of spin 1/2, where $|\psi_+|^2$, $|\psi_-|^2$ are respectively the probabilities at $\mathbf{r}$ of finding the electron with spin s_z parallel and antiparallel to the z direction, then at every point $\mathbf{r}$ there exists a z' direction, depending in general on $\mathbf{r}$, along which the wave function has components $\psi'_+(\mathbf{r}) = 1$, $\psi'_-(\mathbf{r}) = 0$. In other words at each point in space the original wave function represented a polarized electron, the direction of polarization depending in general on $\mathbf{r}$; conversely, although at any point a measurement distinguishing between spin parallel and antiparallel to the z' direction will always find the electron parallel to z', there are finite probabilities, depending on the angle between z and z', of finding the electron spin parallel or antiparallel to any other direction z.

Ensemble of wave functions. The preceding paragraph shows that an unpolarized beam of electrons cannot be represented by a wave function. An unpolarized beam can be thought of as an ensemble of electrons, each of which is polarized in a different direction; at each point $\mathbf{r}$ the direction of polarization depends on the relative phase of the complex quantities $\psi_+(\mathbf{r})$, $\psi_-(\mathbf{r})$. It is concluded, therefore, that many systems are described, not by a wave function, but rather, by an ensemble of wave functions, in which the relative phase of any two projections ψ_m, ψ_n is random when $m \neq n$. Letting the bar denote ensemble average, it follows that (i) $\bar{\rho}_{mn} = 0$ when $m \neq n$; (ii) the ensemble averaged expectation value of any operator A is $\bar{\rho}_{11}A_{11} + \bar{\rho}_{22}A_{22} + \cdots$, where A_{11} is the expectation value of A in the state u_1, and so forth, and $\bar{\rho}_{ii}$ is the probability of finding the system in state u_i. These results are important because they show that in a mixed or impure state characterized by a diagonal density matrix, the expectation value of A is averaged over the populations of states omitting the interference terms characteristically present when the system is describable by a wave function; for instance, the total radiation from a gaseous discharge is the same as if the radiating levels had relative populations $\bar{\rho}_{ii}$ (proportional to the Boltzmann factor when in thermal equilibrium) and radiated independently. *See* STATISTICAL MECHANICS.

[E.G.]

Bibliography: *See* QUANTUM THEORY, NONRELATIVISTIC.

Transformer

An electrical component used to transfer electric energy from one alternating-current (ac) circuit to another by magnetic coupling. Essentially, it consists of two or more multiturn coils of wire placed in close proximity to cause the magnetic field of one to link the other. In general, the transformer will accomplish one or more of the following between two circuits: (1) a difference in voltage magnitude, (2) a difference in current magnitude, (3) a difference in phase angle, (4) a difference in impedance level, and (5) a difference in voltage insulation level, either between the two circuits or to ground.

Transformers are used to meet a wide range of requirements. Pole-type distribution transformers supply relatively small amounts of power to residences. Power transformers are used at generating stations to step up the generated voltage to high levels for transmission. The transmission voltages are then stepped down by transformers at the substations for local distribution. [illegible]

[illegible] band of frequencies. Radio-frequency transformers transfer energy in narrow frequency bands from one circuit to another.

Transformers are often classified according to the frequency for which they are designed. Power transformers are for power-frequency circuits, au-

on the refraction of light at different concentrations, as the optical method for studying the boundary movement.

In all work with the moving boundary methods it is necessary to make corrections for the volume changes at one or the other of the electrodes. These corrections are especially important with the differential method.

Potentials of concentration cells. A composite galvanic cell of the type without transference:

$$\mathrm{Ag};\ \mathrm{AgCl},\ \mathrm{HCl}(C_1),\ \mathrm{H_2(Pt)}:(\mathrm{Pt})\mathrm{H_2},\ \mathrm{HCl}(C_2)\ \mathrm{AgCl};\ \mathrm{Ag}$$

in operation will result in the transport of 1 mole of HCl from the concentration C_1 to the smaller concentration C_2 per faraday F of current passed, and the Gibbs free energy of the process will be EF in which E is the potential of the cell. On the other hand, the operation of the cell with transference

$$\mathrm{Ag};\ \mathrm{AgCl},\ \mathrm{HCl}(C_1):\mathrm{HCl}(C_2),\ \mathrm{AgCl};\ \mathrm{Ag}$$

in which there is a liquid junction between the HCl solutions at the concentrations C_1 and C_2, results in the transport of t_{H^+} equivalent of HCl, per equivalent, t_{H^+} being the Hittorf transference number of the hydrogen ion constituent. The free energy of this process is E_tF in which E_t is the potential of the cell with transference. Thus the transference number $t_{H^+} = E_t/E$. This is true only if the transference number is a constant in the concentration range C_1 to C_2. If it is not, the transference number is given by $t_{H^+} = dE_t/dE$ which is the slope, at any particular concentration, of a plot of E_t against E. These quantities may be expressed as empirical functions and differentiated, or graphical methods may be used. The procedure has been used by MacInnes and J. A. Beattie, by H. S. Harned and E. C. Dreby, and others. This method has not been greatly used for obtaining transference numbers, largely because concentration cells without transference, and with cations other than hydrogen, usually require the use of amalgam electrodes, which involve great experimental difficulties.

However, concentrations cells with transference, together with transference numbers, have been used to obtain activity coefficients of electrolytes. For chlorides, for instance, such cells have the form.

$$\mathrm{Ag};\ \mathrm{AgCl},\ M\,\mathrm{Cl}\ (C_1):M\,\mathrm{Cl}\ (C_2),\ \mathrm{AgCl},\ \mathrm{Ag}$$

in which M is any univalent cation. Such cells have remarkably constant and reproducible potentials, and require only one type of reversible electrode. The thermodynamic equation for the potential E_t of such cells is

$$E_tF = 2RT\int_{C_1}^{C_2} t\, d\ln Cf \qquad (7)$$

in which F is the faraday, R the gas constant, t is the transference number of the cation at the concentration C, ln indicates natural logarithm, and f is the activity coefficient. Much precision data has been obtained in this manner from the work of Longsworth, T. Shedlovsky, and MacInnes on transference numbers and on the corresponding concentration cells. This work has been extended to chlorides with higher valence cations. For the interpretation of the data, Eq. (7) must be appropriately modified. See ELECTRODE POTENTIAL; ELECTROLYSIS; ELECTROMOTIVE FORCE (CELLS).

[D.A.M.]

Bibliography: H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions*, 3d ed., 1958; D. A. MacInnes and L. G. Longsworth, Transference numbers by the method of moving boundaries, *Chem. Rev.*, 11:171, 1932; R. A. Robinson and R. H. Stokes, *Electrolyte Solutions*, 1955.

Transformation theory, quantum

The study of coordinate and other transformations in quantum mechanics which leave invariant fundamental formulas. Matrix notation is convenient for the purposes of the present article and is used throughout. See MATRIX MECHANICS; MATRIX THEORY.

Quantum transformation theory can be interpreted as the theory of rotations and reflections in an infinite-dimensional complex vector space, so-called Hilbert space. (For a discussion of Hilbert spaces, see GEOMETRY, RIEMANNIAN.) In three-dimensional physical space the square of any vector $\mathbf{v}$ is $v^2 = v_1^2 + v_2^2 + v_3^2 = \tilde{v}v$, where v is a column matrix (v_1, v_2, v_3) and $\tilde{v}$ is the row (v_1, v_2, v_3). The transpose $\tilde{A}$ of any matrix A is the new matrix obtained by interchanging rows and columns of A; $(\widetilde{AB}) = \tilde{B}\tilde{A}$. By definition a rotation or reflection of points in space, keeping the coordinate axes fixed, moves the point $\mathbf{v}$ to a new location $\mathbf{v}'$ such that $\tilde{v}'v' = \tilde{v}v$; the new components v' are linear combinations of the old, that is, $v' = Sv$, where S is a three-rowed square matrix. The same transformation $v' = Sv$ of the components could have been achieved by holding points in space fixed and appropriately rotating or reflecting the orthogonal unit vectors (the basis) $\mathbf{i}_1$, $\mathbf{i}_2$, $\mathbf{i}_3$ to new positions $\mathbf{i}_1'$, $\mathbf{i}_2'$, $\mathbf{i}_3'$. In either event the transformation is a rotation or reflection if and only if S is orthogonal, $\tilde{S}S = S\tilde{S} = I$, I being the unit matrix; this requirement guarantees $\tilde{v}'v' = \tilde{v}\tilde{S}Sv = \tilde{v}v$.

With complex vectors the foregoing results remain true, provided the transpose is replaced everywhere by the complex conjugate transpose or adjoint. Thus in quantum theory where $\sum_n \psi_n^*\psi_n = \psi^\dagger\psi = 1$ must hold for a normalized wave function $\psi = \sum_n \psi_n u_n$ expanded in terms of any complete orthonormal set of functions u_n: (i) the set u_n form an orthonormal basis in Hilbert space, (ii) each ψ is a vector in this space, with projections ψ_n on u_n; (iii) expanding in terms of another orthonormal set $\psi = \sum_n \psi_n' u_n'$ corresponds to a rotation or reflection of the basis. Moreover, in such changes of basis: (i) $\psi' = S\psi$, where S is unitary, $S^\dagger S = SS^\dagger = I$; (ii) the new matrix A' representing an operator must be related to the old matrix A by $A' = SAS^\dagger$, in order to ensure invariance of the physically measurable expectation value of A, $\langle A\rangle = \psi^\dagger A\psi$; (iii) given any Hermitian operator $A = A^\dagger$, a basis

The conductors are insulated with special paper or cotton covering, with enamel, or with a combination of both. Large outdoor transformers are immersed in oils to obtain good electrical insulation within small spacings and to provide a cooling medium When lightweight or noninflammable materials are important, transformers may be made with ... cooling me- ... dielec- ... e trans- ... ger for former and through a gas-to-air ... cooling

The low-voltage (LV) winding is usually in the form of a cylinder next to the core. The high-voltage (HV) winding, also cylindrical, surrounds the LV winding as in Fig 2*a*. These windings are often described as concentric windings. The number of turns N may be obtained from the relation

$$E = \frac{fBAN}{22{,}500}$$

where E is the rms voltage, f is the frequency in cps, B is the maximum flux density in kilolines/in ^{2} and A is the cross sectional area of the iron core in square inches

Some manufacturers use a winding arrangement having coils adjacent to each other along the core leg as in Fig 2*b* The coils are wound in the form of a disk with a group of disks for the LV winding stacked alternately with a group of disks for the HV windings. This construction is referred to as interleaved windings

The core sheets are stacked sheet by sheet to form the desired cross sectional area The closed magnetic circuit typically has joints between adjacent sheets, but cores of moderate cross section may be made with a long continuous sheet which has been coiled up to give the required cross section Passages may be provided between groups of sheets for circulation of the cooling oil

For single-phase transformers, the HV and LV coils may be on one leg of a core, with the return path in one two or more other legs The total area of the return legs is equal to that of the main leg An alternative construction has two legs, each with half of the primary windings and half of the secondary windings

A typical three-phase core has three legs, with the HV and LV windings for one phase on each leg The yokes of the core connect between the two outer legs and the middle leg on top and bottom. This core-type construction is shown in Fig 5*a* Another construction sometimes used (shell-type) has the iron as shown in Fig 5*b* Either concentric windings or interleaved windings may be used with either core.

The core and coils are placed in a steel tank with openings for the electrical connections to the windings and for the cooling equipment

Cooling. Small transformers are self-cooled Radiation conduction, and convection from the tank or from radiating surfaces remove the heat generated by the power losses of the transformer On

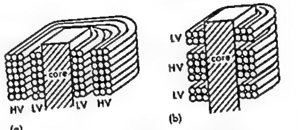


Fig 2 Winding arrangements. (a) Concentric. (b) Interleaved

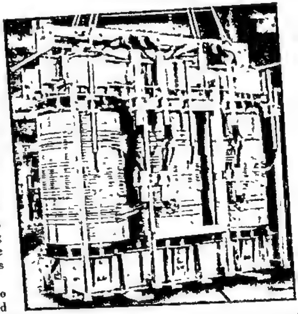


Fig 3 Three-phase transformer core and coils, rated at 50,000 kva, 115,000 volts.

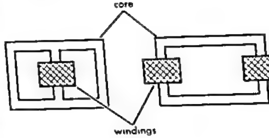


Fig 4 Typical single-phase cores showing location windings.

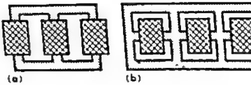


Fig 5. Typical three-phase cores showing locatio windings, (a) Core-type (b) Shell-type

dio transformers for audio-frequency circuits, and so forth. Of course, many of the basic principles of operation apply to all.

POWER TRANSFORMERS

Principle of operation. A power transformer consists of two or more multiturn coils wound on a laminated iron core. At least one of these coils serves as the primary winding. This primary, when connected to an alternating voltage, produces an alternating flux in the core. The flux generates a primary electromotive force, which is essentially equal and opposite to the voltage supplied to it. It also generates a voltage in the other coil or coils, one of which is called a secondary. This voltage generated in the secondary will supply alternating current to a circuit connected to the terminals of the secondary winding. A current in the secondary winding requires an additional current in the primary. The primary current is essentially self-regulated to meet the power (or volt-ampere) demand of the load connected to the secondary terminals. Thus in normal operation energy (or volt-amperes) can be transferred from the primary to the secondary electromagnetically.

Figure 1 shows a transformer with a primary of N_1 turns and a secondary of N_2 turns. A primary voltage V_1 causes a current I_1 to flow through the coil. All quantities shown are alternating, so the arrows indicate only instantaneous polarities.

The magnetic flux ϕ set up by the primary consists of two components. One part passes completely around the magnetic circuit defined by the iron core, thus linking the secondary coil. This is the mutual flux ϕ_m. The second part is a smaller component of flux that links only the primary coil. This is the primary leakage flux ϕ_{l1}. If the secondary circuit is completed through a load, a secondary current I_2 flows and in turn creates a secondary leakage flux ϕ_{l2}. These leakage fluxes contribute to the impedance of the transformer. If the leakage flux is small, the coupling between primary and secondary is said to be close. The use of an iron core decreases the leakage flux by providing a low-reluctance path for the flux. See COUPLED CIRCUITS, MAGNETIC CIRCUITS.

In a power transformer the voltage drops due to winding resistance and leakage are small, therefore V_1 and V_2 are essentially in phase (or 180° out of phase, depending on the choice of polarity). Since the no-load current is small, I_1 and I_2 are essentially in phase (or 180° out of phase). Therefore,

$$V_1 I_1 \simeq V_2 I_2$$

and the voltage ratio

$$\frac{V_1}{V_2} \simeq a$$

where a is the transformation ratio. Substituting the second equation into the first demonstrates that

$$\frac{I_1}{I_2} \simeq \frac{1}{a}$$

the current ratio is inversely proportional to the transformation ratio. A transformer, therefore, may be used to step up or down a voltage from a level V_1 to a level V_2 according to the transformation ratio a. Simultaneously the current will be transformed inversely proportional to a.

The first power equation may be rewritten algebraically as

$$I_1^2 \frac{V_1}{I_1} \simeq I_2^2 \frac{V_2}{I_2}$$

Since V_2/I_2 is the impedance Z_2 of the load on the secondary and V_1/I_1 is the impedance Z_1 of the load as measured on the primary,

$$I_1^2 Z_1 \simeq I_2^2 Z_2$$

or

$$\frac{Z_1}{Z_2} \simeq \left(\frac{I_2}{I_1}\right)^2 \simeq a^2$$

The transformer is thus capable of transforming circuit impedance levels according to the square of the transformation ratio. Extensive use is made of this property in telephone, radio, television, and audio systems.

The transmission of power from primary coil to secondary coil is via the magnetic flux. The flux is proportional to the ampere turns in either coil. Since the power in each coil is nearly the same,

$$N_1 I_1 \simeq N_2 I_2$$

and

$$\frac{N_1}{N_2} \simeq \frac{I_2}{I_1} \simeq a$$

The transformation ratio is therefore approximately equal to the turns ratio.

Construction. Transformer cores are made of special alloy steels rolled to approximately 0.014 in. thick. These thin sheets, or laminations, are stacked to form the transformer core, each sheet being insulated from the others to reduce unwanted eddy-current loss. The steel is heat treated to obtain low hysteresis loss, low exciting current, and low sound level. See CORE LOSS, EDDY CURRENT, HYSTERESIS, MAGNETIC.

Copper conductors are used almost universally. Conductors are round wires in smaller transformers, and rectangular wires (for instance $\frac{1}{2}$ by $\frac{1}{8}$ in.) for larger ones.

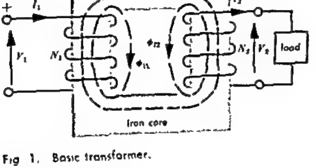


Fig. 1. Basic transformer.

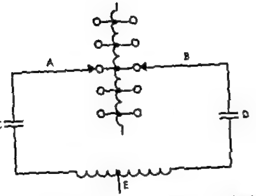


Fig 6 Typical circuit for tap-changing under load

vided half-and half through A and B, has changed, first to be all in B, then partly in A and partly in B,

voltage is required under changing loads It is frequently applied with a tap range of plus or minus 10% of rated voltage It may be made to operate automatically, maintaining a specified voltage at a predetermined point remote from the transformer

Tap-changing under load equipment is used on power transformers supplying residential loads, where variations in voltage would adversely affect the use of lights and appliances. It is also used for chemical and industrial processes, such as on pot lines for the manufacture of aluminum

Parallel operation. Two transformers may be operated *in parallel* (primaries connected to the same source and secondaries connected to the same load) if their turns ratios and per unit impedances are essentially equal A slight difference in turns ratio would cause a relatively large out of-phase circulating current between the two units and result in power losses and possible *overheating*

Phase transformation. Polyphase power may be changed from 3-phase to 6-phase, 3-phase to 12-phase

is of

the g

dc v

cuit

Overloads. Transformers have a *capacity for loading* above their rating Such factors as low ambient temperature and type of load carried may be used to increase the continuous load possible on a given transformer In emergencies it is possible to increase the load further for short times with a *calculable* loss of transformer life Such a load would permit, for instance, a 50% overload for 2 hours following full load [J.H S.]

AUDIO- AND RADIO-FREQUENCY TRANSFORMERS

Audio or video (broad-band) transformers are used to transfer complex signals containing energy at a large number of frequencies from one circuit

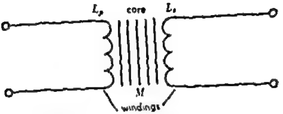


Fig. 7. Schematic of a transformer with symbols

to another. Radio-frequency (rf) and intermediate-frequency (i-f) transformers are used to transfer energy in narrow frequency bands from one circuit to another. Audio and video transformers are required to respond uniformly to signal voltages over a frequency range three to five or more decades wide (for example, from 10 to 100 000 cps), and consequently must be designed so that very nearly all of the magnetic flux threading through one coil also passes through the other. These units are designed to have a coupling coefficient k nearly equal to one

$$k = M/\sqrt{L_pL_s}$$

where L_p and L_s are the primary and secondary inductances respectively, and M is the mutual inductance (Fig 7). The high coupling coefficient is obtained by the use of interleaved windings and a high-permeability iron core, which concentrates the flux. Typical values of k for best quality video transformers may be as high as 0 9998, *that for* power transformers need not be greater than 0.98.

The rf and i-f transformers are built from individual inductors whose magnetic fields are loosely coupled together, $k < 0.30$, and each inductor is resonated with a capacitor to make efficient energy transfer possible near the resonant frequency. *See* RESONANCE (ALTERNATING-CURRENT CIRCUITS).

Audio and video transformers. Audio and video transformers have two resonances (caused by existing stray capacitances) just like many tuned transformers One resonance point is near the low-signal-frequency limit; the other is near the high limit. As the coefficient of coupling in a transformer is reduced appreciably below unity by removal of core material and separation of the windings, tuning capacitors are added to provide efficient transfer of energy The two resonant frequencies combine to one when the coupling is reduced to the value known as critical coupling, then stay relatively fixed as the coupling is further reduced.

All transformers are devices for transferring energy from one circuit to another. The energy transferred is absorbed either in the circuits themselves or in an external load circuit. For this reason, proper termination is essential to optimum behav-

minimum

at which the open-circuit reactance of the primary is approximately twice its

larger units, fans are sometimes added to the radiating surfaces. A transformer may have one rating with a basic method of cooling and a higher rating with supplemental cooling. Pumps may be added to give further cooling. An oil-to-air heat exchanger having finned tubes is used on the very large units. This equipment has a pump for circulating oil, and fans for forcing the air against the heat exchanger. Water cooling may be used with cooling coils or with an oil-to-water heat exchanger having an oil pump.

Characteristics. The service conditions for a particular transformer are considered by the designer in choosing materials and the arrangement of parts. The final design then may be measured by test with respect to a number of characteristics.

No-load loss. The sum of the hysteresis and eddy loss in the iron core is the no-load loss.

Exciting Current. The exciting current is that supplied to the transformer at no load when operating at rated voltage. This current energizes the core and supplies the no-load loss. Due to the characteristic shape of the *B-H* curve of iron, the current is not a true sine wave, but has higher frequency harmonics. In a typical power transformer the exciting current is so small (usually less than 1%) that I_2 is approximately $(N_1/N_2)I_1$. In this sense the ampere-turns in the two windings are said to balance.

Load Loss. This is the sum of the copper loss, due to the resistance of the windings (I^2R loss), plus the eddy-current loss in the winding, plus the stray loss (loss due to flux in metallic parts of the transformer adjacent to the windings, the flux resulting from current in the windings).

Total loss and efficiency. The total loss in a transformer is the sum of the no-load and full-load losses. Representative values for a 20,000-kva, three-phase, 115-kv power transformer are no-load loss, 42 kw, load loss, 85 kw; and total loss, 127 kw.

$$\text{Efficiency} = \frac{\text{output in kw}}{\text{input in kw}} = \frac{\text{output}}{\text{output} + \text{losses}}$$

For this transformer the efficiency is 20,000/20,127 or 99.37%.

Voltage ratio. This is the ratio of voltage on one winding to the voltage on another winding at no load. This is the same as the turns ratio.

Impedance. Consider a transformer having equal turns in the primary and secondary windings. If one side is connected to a generator and the other side to a typical power system load, the voltage measured on the load side will be less than that on the generator side, by the amount of the impedance drop through the transformer.

Impedance is measured by connecting the secondary terminals together (short-circuited) and applying sufficient voltage to the primary terminals to cause rated current to flow in the primary winding. The transformer impedance in ohms equals the primary voltage divided by the primary current. Impedance is usually referred to the transformer kva and kv base and given as per cent impedance. Per cent reactance is usually close in value to per cent impedance, since the per cent resistance is small.

$$\%\ \text{impedance} = \frac{1}{10}\frac{\text{kva}}{(\text{kv})^2} \times \text{ohms}$$

$$\%\ \text{resistance} = \frac{\text{load loss in kva}}{\text{kva rating}} \times 100$$

Typical values for a 20,000-kva, three-phase, 115-kv self-cooled power transformer are resistance, 0.4% and impedance, 7.5%.

Regulation. Regulation is the change in output (secondary) voltage that occurs when the load is reduced from rated value to zero, with the primary impressed terminal voltage maintained constant. This is usually expressed as a per cent of rated output voltage at full load (E_{FL}).

$$\%\ \text{regulation} = \frac{E_{NL} - E_{FL}}{E_{FL}} \times 100$$

where E_{NL} is the output voltage at no load. When a transformer supplies a capacitive load, the power factor may cause a higher full-load voltage than no-load voltage.

Cooling. Temperature tests (heat run tests) are made by operating the transformer with total losses until the temperatures are constant. In the United States the standard winding rise is 55°C over a 30°C air ambient.

Insulation. Sufficient insulation strength must be built into a transformer to withstand normal operation at its rated voltage, and system voltage transients due to lightning and switching surges.

Audio sound. The iron core lengthens and shortens due to magnetostriction during each voltage cycle, giving rise to a hum having a frequency twice that of the voltage. This and other frequencies may cause mechanical vibrations in different parts of the transformer due to resonance.

Taps. The application of a transformer to a power system involves a correct choice of turns ratio for average operating conditions, and the selection of proper taps to obtain improved voltage levels when average conditions do not prevail.

Tap-changers are frequently used in the HV winding to give plus or minus two 2½% taps (5% above and 5% below rated voltage). These taps may be changed only when the transformer is deenergized, that is, when the service is interrupted.

Tap-changing under load. A special motor-driven tap-changer is used to permit tap-changing when the transformer is energized and carrying full load. One of its simpler forms is shown in Fig. 6. The transformer taps are brought to a tap-changer having two sets of fingers, A and B. Initially these are on the same tap. When a change is required, a contactor C opens, and A moves to the next lower tap. C now closes. Next D opens and B moves down to the same tap as A. The current, which initially di-

lightning arrester applications for stations and substations, *Trans. AIEE*, 76(3): 614–627, 1957; American Standards Association, ASA C57.12, 1949, C57.12a, 1954; L. F. Blume (ed.), *Transformer Engineering*, 2d ed., 1951; A. N. Garin, Zero-phase-sequence characteristics of transformers, *Gen. Elec. Rev.*, 43:131–136, 174–179, 1940; E. M. Hunter, J. R. Meador, and W. J. Rudge, Reduced transformer insulation, *Trans. AIEE*, 76(3): 34–38, 1957; Mass. Inst. Technol., *Magnetic Circuits and Transformers*, 1943; A. F. Puchstein, T. C. Lloyd, and A. G. Conrad, *Alternating Current Machines*, 3d ed., 1954; K. A. Pullen, *Conductance Design of Active Circuits*, 1959.

Transfusion

The administration of blood, or one of its components, as a part of treatment. The techniques and purposes of transfusion are often similar to those of other fluids, and therefore the use of such substances as glucose, synthetic plasma expanders, and related agents may be considered in addition to blood.

There are certain fairly well delineated indications for the use of some form of transfusion. Hemorrhage, severe burns and certain forms of shock are perhaps the most important conditions for which blood transfusion is utilized. Other disorders in which hemotherapy may be indicated include hemophilia, leukemia, certain anemias, and the rather rare hereditary or familial metabolic disorders in which some portion of the blood is lacking or deficient. See HEMATOLOGIC DISORDERS.

Blood may be administered in several forms, de-

... which there is a suitable anticoagulant present. Various blood portions or derivatives are prepared by appropriate methods. Their advantage lies in the fact that such preparations may be stored and stockpiled for much greater periods of time than is currently possible with whole blood. Packed red blood cells, dried or fluid blood plasma, and plasma protein preparations such as albumin or globulin are examples in most common use.

Standards. Certain standards have been established for the various aspects of blood donation, collection, testing and typing, storage, and administration to a recipient. In the United States, the many agencies involved in these activities have developed appropriate criteria for each of the phases mentioned. The American Association of Blood Banks and the Joint Blood Council publish such standards for associated hospital and other medical groups, while various government agencies, such as the Defense Department, perform similar coordinating functions for civil and military federal units. The National Institutes of Health provide technical standards for cross-matching tests, as well as numerous other facilities. Many other agencies, operating on national, state, or local levels, cooperate in providing transfusion services.

Donor. A donor should be an adult in good health and free of disease transmissible by the blood. It has been found that 450 ml of blood is an optimum quantity to obtain from the donor at intervals which, although varying widely in relation to other factors, should not be more frequent than once in 6–8 weeks. A history of viral hepatitis or recent exposure to it, recent pregnancy, or a record of severe allergic disorder will eliminate a prospective donor. In addition, the donor must not be anemic, should not have had dental extraction or surgery in the few days prior to collection, and should not have received a transfusion in the preceding several months.

Blood collection. Collection of blood follows rigid standards so that the specimens obtained are not contaminated by bacteria or pyrogenic substances. Although the physical details of collection vary from one agency to another, due regard is given to the primary prerequisite mentioned above, as well as physical and psychic welfare of the donor.

Whole blood is collected in sterile, pyrogen-free containers of various types, all of which must meet rigid self-imposed standards. During collection, usually by means of a disposable plastic intravenous set, the blood is added to an adequate amount of anticoagulant already present in the container.

Typing. Each single sample is usually typed twice and by different methods, so that error is minimal. In addition, if whole blood is transported to another location for storage, it is again typed to check on the accuracy of the initial classification.

The typing is principally concerned with the ABO system, and secondarily, but no less important, with the Rh system. Although other groups of antibodies may be present, these two sets are of the most practical use when dealing with large numbers of persons. Other supplementary tests, or check tests, are run as indicated. See BLOOD GROUPS.

Storage. Whole blood is commonly stored at 1–6°C (34–43°F) for various lengths of time, depending on the methods and materials employed in its collection. While most facilities retain such blood for 1–2 months, experimental work shows great promise in extending this to a much longer, or even indefinite time.

Administration. When blood is to be given, a sample of the recipient's blood is subjected to compatibility tests with the blood of the donor. It is necessary to determine whether the donor cells will be damaged, or hemolyzed, by the recipient's plasma, or less important, whether the recipient's cells will be harmed by the donor's plasma. If such compatibility tests are negative, the transfusion proceeds; if they are positive, either another specimen is used or appropriate safeguards are employed if no other blood is available and ...

fu

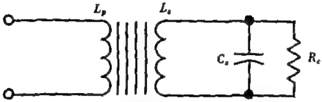


Fig. 8 Circuit of loaded transformer.

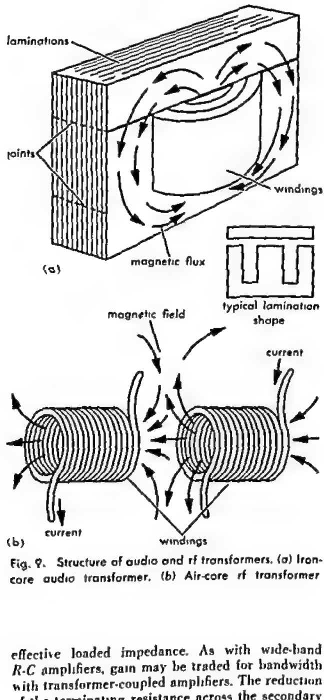


Fig. 9. Structure of audio and rf transformers. (a) Iron-core audio transformer. (b) Air-core rf transformer

effective loaded impedance. As with wide-band R-C amplifiers, gain may be traded for bandwidth with transformer-coupled amplifiers. The reduction of the terminating resistance across the secondary of the transformer reduces the minimum operating frequency f_1 and, in the presence of output capacitance, it raises the maximum frequency f_2. The approximate values of the minimum and maximum frequencies and the resonant frequency f_r are given by

$$f_1 = R_c(N_p/N_s)^2/\pi L_p$$
$$f_2 = 1/2\pi R_c C_s$$
$$f_r = 1/2\pi\sqrt{L_{ss}C_s}$$

where $L_{ss} = L_s - M^2/L_p = L_s(1 - k^2)$

This L_{ss} is the secondary inductance with the primary short-circuited; C_s is the output capacitance, both external and internal, on the transformer; R_c is the load resistance (see Fig. 8). The resonant frequency f_r should be larger than f_2 for best performance. See AMPLIFIER.

A transformer used to activate terminating circuitry is called an output transformer; one to activate an input circuit is an input transformer; others are called interstage transformers.

Distortion. The distortion introduced into the amplified signal by a transformer is caused primarily by its hysteresis loss. This loss may be minimized by proper loading on the secondary. The load component of current then is large compared to the magnetizing current. In addition, a resistive load keeps the amplification uniform as a function of frequency, and keeps the phase distortion to a minimum.

The magnetic core in an audio or video transformer is subject to two kinds of saturation, that due to applied direct current in the windings, and that due to excessively large signal currents. The direct current in the windings may make the hysteresis loop of the iron core nonsymmetrical, necessitating the use of a larger core having a built-in airgap Both large signal amplitudes and low frequencies can cause signal saturation to occur in the core. The structure of audio and rf transformers is shown in Fig 9.

Rf and i-f transformers. These use two or more inductors, loosely coupled together, to limit the band of operating frequencies Efficient transfer of energy is obtained by resonating one or more of the inductors By using higher than critical coupling, a wider bandwidth than that from the individual tuned circuits is obtained, while the attenuation of side frequencies is as rapid as with the individual circuits isolated from one another.

The tuning of the primary, the secondary, or both may be accomplished either by the variation of the tuning capacitor or by an adjustable magnetic or conducting slug that varies the inductance of the inductor (see Fig 10)

The operating impedance of a tuned circuit of an rf transformer is a function of its Q and its tuning capacitance In general, high-power circuits require a high capacitance for energy storage and therefore have low values of impedance In any application, the impedance level must be kept sufficiently small to prevent instability and oscillation. See RADIO FREQUENCY AMPLIFIER [K.A.P.]

Bibliography: AIEE Working Group, Report on

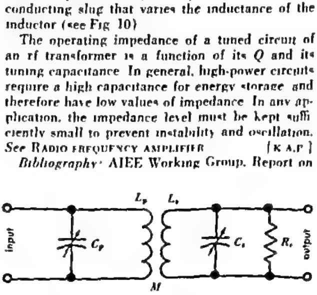


Fig. 10. Tuned rf transformer.

II The sum of the voltages e around a mesh is equal to zero: $\Sigma e = 0$

The voltage drops associated with a resistance, inductance, mutual inductance, and capacitance are, respectively,

$$e_R = Ri \qquad e_L = L\frac{di}{dt}$$
$$e_{12} = M_{12}\frac{di_2}{dt} \qquad e_C = \frac{1}{C}\int i\,dt \tag{3}$$

(see COUPLED CIRCUITS). In applying Kirchhoff's law to a closed mesh it is merely necessary (with due regard for signs and polarities) to equate the sum of all the voltage sources (such as generators and batteries) to the sum of all the voltage drops in the mesh, thus

$$\sum_m e_m = \sum_m \sum_n \left(R_{mn} i_n + L_{mn}\frac{di_n}{dt} + \frac{1}{C_{mn}}\int i_n\,dt \right) \tag{4}$$

in which m refers to any branch of the mesh in question, and n to any current causing a voltage drop *in that branch* (a branch may carry currents from meshes other than the mesh [illegible]

mutu [illegible]

(4) [illegible]

cuit [illegible]

sometimes convenient to make use of charges q and fluxes ϕ by the substitutions:

$$i = \frac{dq}{dt} \quad \text{or} \quad q = \int i\,dt \quad \text{and} \quad N\phi = Li \tag{5}$$

Since Eq. (4) contains an integral, it is convenient to eliminate it either by differentiating once, or by substituting Eq. (5). Then

$$\sum_m \frac{de_m}{dt} = \sum_m \sum_n \left(R_{mn}\frac{di_n}{dt} + L_{mn}\frac{d^2 i_n}{dt^2} + \frac{i_n}{C_{mn}} \right) \tag{6}$$

$$\sum_m e_m = \sum_m \sum_n \left(R_{mn}\frac{dq_n}{dt} + L_{mn}\frac{d^2 q_n}{dt^2} + \frac{q_n}{C_{mn}} \right) \tag{7}$$

It is customary to factor out the current from Eq. (4) and write the equation in terms of a generalized impedance operator Z_{mn} defined by

$$\sum_m e_m = \sum_m \sum_n \left(R_{mn} + L_{mn}\frac{d}{dt} + \frac{1}{C_{mn}}\int dt \right) i_n = \sum_m \sum_n Z_{mn} i_n \tag{8}$$

Since an equation of this type can be written for each of the N meshes of the network, the totality of the differential equations for the entire network constitutes a system of simultaneous differential equations of the form

$$\begin{aligned} e_1 &= Z_{11}i_1 + Z_{12}i_2 + \cdots + Z_{1N}i_N \\ e_2 &= Z_{21}i_1 + Z_{22}i_2 + \cdots + Z_{2N}i_N \\ &\cdots \\ e_N &= Z_{N1}i_1 + Z_{N2}i_2 + \cdots + Z_{NN}i_N \end{aligned} \tag{9}$$

In general, the process of solution of such a set of equations for any current leads to a differential equation of order $2N$ and there will be $2N$ integration constants associated with it, and these *integration constants* must be determined from the initial conditions at the first instant of the disturbance.

The solution of an ordinary differential equation with constant coefficients is in two parts:

$$\begin{pmatrix}\text{Complete}\\ \text{solution}\end{pmatrix} = \begin{pmatrix}\text{particular}\\ \text{integral}\end{pmatrix} + \begin{pmatrix}\text{complementary}\\ \text{solution}\end{pmatrix} \tag{10}$$

The particular integral depends on the form of the applied voltage, and represents the final steady-state solution.

The complementary solution is independent of the form of the applied voltages, but depends on *the initial conditions in the circuit*, and represents the transient terms. Thus Eq. (10) is the mathematical equivalent of Eq. (1).

The opening of a switch may be simulated by superimposing a current equal and opposite and thereby canceling the current through the switch. Likewise the closing of a [illegible]

lation current (or voltage) may be superimposed on the currents and voltages already existing in the system, *that is:*

$$\begin{pmatrix}\text{Resultant}\\ \text{voltages}\\ \text{and cur-}\\ \text{rents}\end{pmatrix} = \begin{pmatrix}\text{those which}\\ \text{would exist if}\\ \text{switches were}\\ \text{not operated}\end{pmatrix} + \begin{pmatrix}\text{those due to}\\ \text{the "cancel-}\\ \text{lation" volt-}\\ \text{ages or cur-}\\ \text{rents}\end{pmatrix}$$

In many practical cases transients are one of three types:

1. Single-energy transients, in which only one form of energy storage (either *electromagnetic* or *electrostatic*) is present, and the transient exhibits simple exponential decay from the initial to the final conditions.

2. Double-energy transients, in which both forms of energy storage are present, and the transient is either aperiodic or a damped sinusoid.

3. *Combination of 1 and 2.*

DC TRANSIENTS

Transients are initiated in dc circuits either when a switch is closed on a dc voltage, or when a switch is opened and the detached circuit is permitted to discharge its energy sources. The assumption is made that the suddenly applied dc voltage is sustained at a constant value for an indefinite period after the switching operation, and time is counted from the instant, $t = 0$, at which the switch is operated. It is further usually assumed, where possible, that all *circuit parameters* (resistances R, inductances L, and capacitances C) are constant, so that the resulting differential equations are linear.

The energy sources *in* dc systems comprise dynamoelectric machines, *mercury-arc* rectifiers, vacuum tubes, and electric batteries. Energy W is stored in a circuit *in inductances* ($W = \frac{1}{2}Li^2$) and capacitances ($W = \frac{1}{2}Ce^2$). Energy is dissipated in circuit resistances ($W = Ri^2$).

contains a filter system to prevent the passage of any aggregation or coagulum of the donor's blood. Such units, readily employed with the storage container, have recently been made of plastic and are disposable.

The actual transfusion is watched carefully to maintain adequate technique and to be ready for the reaction which will rarely appear, despite the multiple prior testing. Transfusion reactions are of two general types, those due to pyrogenic activity and those due to antigenic incompatibility of the blood so that red cell destruction, or hemolysis, occurs. Clinically, the reactions may be difficult to distinguish. The reaction may be quite transient and mild, or rarely, may result in a true serum sickness with death ensuing. Although such severe reactions are statistically remarkably few, they do occur and thus provide a never-ending stimulus for watchfulness on the part of the physician and attendants. The subjective and psychic effects of a transfusion may also be of great significance, particularly in an emotionally labile individual.

Multiple transfusions, either on a single occasion or over a period of time, present further potential complications which are best evaluated and handled by physicians who are expert in blood work.

Research. The continuing and increasing use of blood as a therapeutic agent as well as a necessary component in any stockpiling of essential materials for a disaster area has produced a great number of investigations to develop this field further. Research goals include better methods of storage of whole blood and its derivatives, the development of blood substitutes, the refinement of technical methods and testing, investigation of many basic questions regarding blood function in normal and disease states, and the education, not only of technical personnel, but also of the population at large in matters pertaining to blood transfusions and related methods of therapy. [E.C.ST.]

Transient, electric

A temporary component of current and voltage in an electric circuit which has been disturbed. In ordinary circuit problems, a stabilized condition of the circuit is assumed and steady-state values of current and voltage are sufficient (*see* ALTERNATING-CURRENT CIRCUIT THEORY; DIRECT-CURRENT CIRCUIT THEORY). However, it often becomes important to know what occurs during the transition period following a circuit disturbance until the steady-state condition is reached. Transients occur only in circuits containing inductance or capacitance. In general, transients accompany any change in the amount or form of energy stored in the circuit. Both direct-current (dc) transients and alternating current (ac) transients are treated following the introduction.

Introduction. The study of transient phenomena is very broad. The mathematical requirements become severe and go far beyond the borders of all known mathematics. Transient analysis often requires the use of calculating machines, models, and tests. Fourier and Laplace transforms have proven indispensable in the modern treatment of transients and these disciplines need be mastered by anyone going far in the study of transients. The analysis of lumped-parameter circuits is comparatively easy and is all that will be described here, but transients of a much more complex nature occur on distributed-parameter circuits, such as transmission lines.

An electric circuit or system under steady-state conditions of constant, or cyclic, applied voltages or currents is in a state of equilibrium. However, the circuit conditions of voltage, current, or frequency may change or be disturbed. Also, circuit elements may be switched in or out of the circuit. Any change of circuit condition or circuit elements causes a transient readjustment of voltages and currents from the initial state of equilibrium to the final state of equilibrium. In a sense the transient may be regarded as superimposed on the final steady-state, so that

$$\begin{pmatrix}\text{Instantaneous}\\ \text{condition}\end{pmatrix} = \begin{pmatrix}\text{final}\\ \text{condition}\end{pmatrix} + \begin{pmatrix}\text{transient}\\ \text{terms}\end{pmatrix} \qquad (1)$$

Furthermore, since the instantaneous condition at the first instant of disturbance (time zero) must be the initial condition,

$$\begin{pmatrix}\text{Initial}\\ \text{condition}\end{pmatrix} = \begin{pmatrix}\text{final}\\ \text{condition}\end{pmatrix} + \begin{pmatrix}\text{transient terms}\\ \text{at time zero}\end{pmatrix} \qquad (2)$$

A great deal of information may be obtained from these two "word equations" without recourse to mathematics. For example, if the weight on the end of a vertical spring is suddenly increased, its final displacement can be determined. Since the spring-weight combination is known to be an oscillating system, the amplitude of the transient oscillation follows from Eq. (2) as the difference between the initial and final displacements.

The nature of an electric transient is determined by three things: (1) the circuit or network itself—the interconnections of its elements and the circuit parameters (resistances R, inductances L, capacitances C, mutual inductances M); (2) the initial conditions of voltages, currents, charges, and flux-linkages at the start of the transient; and (3) the nature of the disturbance which initiated the transient.

The circuit, or network, is usually defined by a diagram of connections showing all of the interconnections, junctions, meshes, circuit parameters, voltage and current sources and their polarities, and switches. Corresponding to the network a differential (or integral differential) equation, or a set of such equations, may be written. These equations may also be written in terms of operational calculus, or as Laplace transforms, or in any other suitable mathematical equivalents. The equations are established in accordance with Kirchhoff's laws:

1. The sum of the currents i at a junction is equal to zero: $\Sigma i = 0$

Canceling the constant A and the exponential, there results a quadratic in a whose solution gives two possible values

$$a_1 = \frac{-RC + \sqrt{R^2C^2 - 4LC}}{2LC} \tag{28}$$

and

$$a_2 = \frac{-RC - \sqrt{R^2C^2 - 4LC}}{2LC}$$

Associating the integration constant A_1 with a_1 and A_2 with a_2 the solution takes the form

$$i = A_1\epsilon^{a_1t} + A_2\epsilon^{a_2t} \tag{29}$$

The voltage across the capacitor is, by Eq. (24)

$$e_C = \frac{1}{C}\int i\,dt = E - Ri - L\frac{di}{dt}$$
$$= E - (R + a_1L)A_1\epsilon^{a_1t} - (R + a_2L)A_2\epsilon^{a_2t} \tag{30}$$

Initially at $t = 0$, the current must be zero because of the inductance, and the capacitor voltage is $e_C = V$. By the first of these conditions, $i(0) = 0$ in Eq. (29) it is seen that $A_2 = -A_1$. And by the second condition $e_C(0) = V$, in Eq. (30)

$$V = E - (R + a_1L)A_1 - (R + a_2L)A_2$$
$$= E - (a_1 - a_2)LA_1 \tag{31}$$
$$A_1 = \frac{E - V}{(a_1 - a_2)L} = \frac{C(E - V)}{\sqrt{R^2C^2 - 4LC}}$$

The complete solution then is

$$i = \frac{C(E - V)}{\sqrt{R^2C^2 - 4LC}}(\epsilon^{a_1t} - \epsilon^{a_2t}) \tag{32}$$

There are three special cases of this solution, depending on the nature of the radical in Eq. (32).

Nonoscillatory case: $R^2C > 4L$. In this case the radical is positive, and the exponents a_1 and a_2 are real and negative, $a_1 = -\alpha + \beta$, $a_2 = -\alpha - \beta$, where $\alpha = R/2L$ and $\beta = \sqrt{R^2C^2 - 4LC}/2LC$

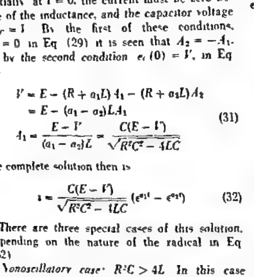


Fig. 3 Resistance, inductance, and capacitance in series

Then

$$i = \frac{C(E - V)}{\sqrt{R^2C^2 - 4LC}}\epsilon^{-\alpha t}(\epsilon^{+\beta t} - \epsilon^{-\beta t}) \tag{33}$$
$$= \frac{2C(E - V)}{\sqrt{R^2C^2 - 4LC}}\epsilon^{-\alpha t}\sinh\beta t$$

This is a double exponential decay and is shown in Fig. 3, where q_0 is assumed to be 0 and $V = 0$.

Aperiodic case: $R^2C = 4L$. In this case $a_1 = a_2 = -R/2L$ and the radical in the denominator of Eq. (32) is zero, as is the numerator. The indeterminant is easily evaluated from Eq. (33) upon letting $\beta \to 0$; thus

$$i = \frac{2C(E - V)}{2LC\beta}\epsilon^{-\alpha t}\sinh\beta t|_{\beta\to 0} = \frac{E - V}{L}\epsilon^{-\alpha t}t \tag{34}$$

This is the aperiodic or critical case, and is illustrated in Fig. 3.

Oscillatory case: $R^2C < 4L$. In this case the radical in Eq. (28) becomes imaginary and the exponents take the form

$$a_1 = \frac{-R}{2L} + j\frac{\sqrt{4LC - R^2C^2}}{2LC} = -\alpha + j\omega$$
$$a_2 = \frac{-R}{2L} - j\frac{\sqrt{4LC - R^2C^2}}{2LC} = -\alpha - j\omega$$

and Eq. (32) becomes

$$i = \frac{C(E - V)}{2j\omega LC}\epsilon^{-\alpha t}(\epsilon^{j\omega t} - \epsilon^{-j\omega t}) = \frac{E - V}{\omega L}\epsilon^{-\alpha t}\sin\omega t \tag{35}$$

This is a damped oscillation and is illustrated in Fig. 3.

AC TRANSIENTS

AC transients differ from dc transients in two im portant respects: (1) the final condition, or stead state is an alternating or cyclic one and (2) th amplitudes of the transient terms depend on th point on the ac applied voltage wave at which th transient is initiated, and can therefore have ma different values or even change sign.

In the section on dc transients, solutions we carried out for RL, RC, and RLC circuits switch onto a voltage source. In the present section, solution will be carried out in detail for the R circuit only, and the others will be regarded special cases of this general solution by putt $C = \infty$ and $L = 0$, respectively.

Resistance, inductance and capacitance series. Consider the circuit of Fig. 4 in which RLC circuit is suddenly switched onto an ac vol

$$e = E\sin(\omega t + \gamma)$$

at an electrical angle γ displaced from ωt. Assume that a current I is flowing in the ci and a voltage V is across the capacitance a instant the switch is closed on the alternator. voltage equation and the initial conditions are

$$E\sin(\omega t + \gamma) = Ri + L\frac{di}{dt} + \frac{1}{C}\int i\,dt$$

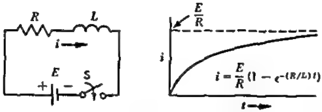


Fig. 1. Resistance and inductance in series

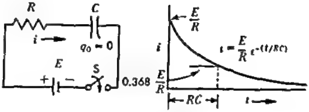


Fig. 2. Resistance and capacitance in series

Resistance and inductance in series. A lumped parameter circuit comprising a dc voltage source E, a resistance R, and an inductance L is shown in Fig. 1. When the switch S is closed, the voltage E is *suddenly impressed on the series circuit.* By Kirchhoff's second law the differential equation is

$$E = Ri + L\frac{di}{dt} \tag{11}$$

Equating the right-hand side of Eq. (11) to zero, the complementary solution is found by assuming

$$i = I\epsilon^{at} \tag{12}$$

in which I and a are constants to be determined. Upon substitution in Eq. (11) there results

$$E = RI\epsilon^{at} + aLI\epsilon^{at} = (R + aL)i \tag{13}$$

from which there follows $a = -R/L$.

The particular integral is the final steady-state solution or $i(\infty) = E/R$. The complete solution is the sum of the complementary solution and the particular integral

$$i = E/R + I\epsilon^{-(R/L)t} \tag{14}$$

But this expression contains the unknown integration constant I which must be found from the initial conditions. At $t = 0$, the instant at which the switch is closed, there can be no current in the *circuit, since it is impossible to store energy in an* inductance instantaneously. Therefore $i(0) = 0$ and when this is put in Eq. (14)

$$i(0) = 0 = E/R + I \quad \text{or} \quad I = -E/R \tag{15}$$

The complete solution by Eq. (14) therefore is

$$i = \frac{E}{R} - \frac{E}{R}\epsilon^{-(R/L)t} \tag{16}$$

which is in the form of Eq. (1), in which E/R *is the final* (steady-state) condition and $-(E/R)\epsilon^{-(R/L)t}$ is the transient term. Note also that at $t = 0$ this equation gives $i(0) = 0$, agreeing with Eq. (2). The graph of the transient is shown in Fig. 1, which shows the current starting at zero, rising exponentially, and approaching its steady-state value of E/R as t approaches infinity.

Resistance and capacitance in series. A resistance-capacitance series circuit is shown in Fig. 2. A residual charge q_0 is on the capacitor just prior to closing the switch S.

The differential equation of the circuit is

$$E = Ri + \frac{1}{C}\int i\,dt \tag{17}$$

This equation may be solved as it is, or converted to a linear differential equation by differentiation

$$0 = R\frac{di}{dt} + \frac{i}{C} \tag{18}$$

or restated in terms of the instantaneous charge on the capacitor by substituting $i = dq/dt$; thus

$$E = R\frac{dq}{dt} + \frac{q}{C} \tag{19}$$

The solution to this equation, following precisely the steps of the previous paragraph, is

$$q = Q\epsilon^{-(t/RC)} + CE \tag{20}$$

in which Q is the integration constant. But initially, at $t = 0$, there was a residual charge q_0 on the capacitance and therefore

$$q(0) = q_0 = Q + CE \quad \text{or} \quad Q = q_0 - CE \tag{21}$$

and upon the substitution of Eq. (21) in Eq. (20), it is seen that

$$q = CE - (CE - q_0)\epsilon^{-(t/RC)} \tag{22}$$

This equation is in the form of Eq. (1), in which CE is the final (steady-state) condition, and the other term is the transient. Initially, at $t = 0$, $q = q_0$ in agreement with Eq. (2). The current is

$$i = \frac{dq}{dt} = \frac{CE - q_0}{RC}\epsilon^{-(t/RC)} = \left(\frac{E}{R} - \frac{q_0}{RC}\right)\epsilon^{-(t/RC)} \tag{23}$$

Eqs. (22) and (23) have been plotted in Fig. 2 for the case $q_0 = 0$.

Resistance, inductance, and capacitance in series. This circuit is shown in Fig. 3. The capacitor is assumed to have an initial charge q_0, or an initial voltage $V = q_0/C$. The differential equation of the circuit is

$$E = Ri + L\frac{di}{dt} + \frac{1}{C}\int i\,dt \tag{24}$$

Differentiating once to clear the integral

$$0 = R\frac{di}{dt} + L\frac{d^2i}{dt^2} + \frac{i}{C} \tag{25}$$

This is a second-order linear differential equation with constant coefficients, and since it is equated to zero, it will possess only a complementary solution. Assume

$$i = A\epsilon^{at} \tag{26}$$

which upon substitution in Eq. (25) yields

$$0 = aRA\epsilon^{at} + a^2LA\epsilon^{at} + (A/C)\epsilon^{at} \tag{27}$$

and the capacitor voltage is, by Eq (44)

$$e_C = \frac{-E}{\omega CZ}\cos(\omega t + \gamma - \theta) + \left[V + \frac{E}{\omega CZ}\cos(\gamma - \theta)\right]\epsilon^{-t/RC} \quad (55)$$

Thus the transient, starting with a voltage V, decays exponentially to its final ac steady-state value. If the switch is closed at an angle γ on the voltage wave such that

$$V = -\frac{E}{\omega CZ}\cos(\gamma - \theta) \quad (56)$$

there will be no transient. [L.V.B.]

Bibliography S Fich, *Transient Analysis in Electrical Engineering*, 2d ed, 1951, H H Skilling, *Transient Electric Currents*, 2d ed, 1952, Charles P. Steinmetz, *Transient Electric Phenomena and Oscillations*, 3d ed, 1920, Ernst Weber, *Linear Transient Analysis*, 2 vols, 1954

Transistor

An active component of an ele...

...tacts are made At least one contact is ohmic (nonrectifying), and at least one contact is rectifying Usually there are two closely spaced rectifying contacts and one ohmic contact.

For a discussion of rectifying contacts, see JUNCTION DIODE; POINT-CONTACT DIODE, SEMICONDUCTOR RECTIFIER

The operation of a simple transistor consists of the control of the current flowing in the high-resistance direction through one rectifying contact (called the collector) by the current flowing in the low resistance direction in the other rectifying contact (called the emitter) The third contact, which is ohmic, is called the base contact.

These contacts usually consist of two or more *regions* The regions in which the actual rectification processes take place are called the emitter barrier and collector barrier The region between these two barriers is called the base region, or simply the base *The* regions outside of these barriers are called the *emitter* and collector regions.

Transistors are used in radio receivers, in electronic computers in electronic instrumentation and *control* equipment and in almost any electronic circuit where vacuum tubes are useful and the required voltages are not too high. Transistors have the advantage over their vacuum-tube counterparts in that they *are much* smaller, consume less power and have no filament to burn out They are at a disadvantage in that they do not yet operate at as high voltages as vacuum tubes and their action is degraded at high temperatures

Classification of transistors. Transistors are classified chiefly by four *criteria: first*, by the type and number of rectifying contacts, second, *by the* technology used in their fabrication, *third*, by the variations in their principle of operation; and fourth, by the semiconductor material used In practice, the designation of a single transistor usually may include several of the above criteria.

One common type of transistor is the *germanium p-n-p* alloy-junction transistor. Here the *p-n-p* stands for the conductivity type of the emitter, base, and collector regions, respectively (the first criterion above). The *n* stands for negative since the charge on an electron is negative and electrons carry most of the current in a region of *n* type conductivity. In a region of p-type conductivity most of the current is carried by electron vacancies, called holes, which behave as if they were positively charged. For a discussion of conductivity type, *see* SEMICONDUCTOR.

The term alloy-junction in this transistor designation refers to the second *criterion. The emitter* and collector regions were fabricated by recrystallization from an alloy of some suitable metal which had previously been fused in contact with the opposite surfaces of the original *n*-type semiconductor body and had dissolved some of the semiconductor material *Fused* junction is equivalent terminology Naming the semiconductor material

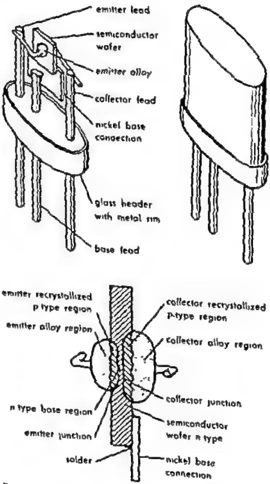


Fig 1 View and section of alloy-junction transistor

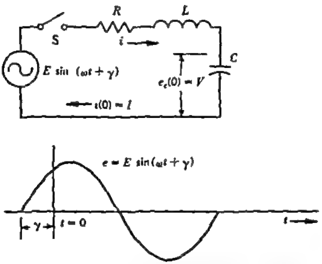


Fig 4 RLC circuit with alternating-current source

$$i = I \quad \text{and} \quad e_C = V \text{ at } t = 0 \tag{38}$$

Differentiating Eq. (37) once to clear it of the integral

$$\omega E \cos(\omega t + \gamma) = L\frac{d^2 i}{dt^2} + R\frac{di}{dt} + \frac{i}{C} \tag{39}$$

The complementary solution of this equation is the same as in the dc case of Eq (25), that is, Eq. (29) where a_1 and a_2 are given in Eq. (28). The particular integral of Eq (39) is its final steady-state current This is most easily obtained as the ordinary ac solution for the current

$$i_{ac} = \frac{E}{Z}\sin(\omega t + \gamma - \theta) \tag{40}$$

in which
$$Z = \sqrt{R^2 + (\omega L - 1/\omega C)^2} \tag{41}$$

$$\tan\theta = \frac{\omega L - 1/\omega C}{R} = \frac{\omega^2 LC - 1}{\omega CR} \tag{42}$$

The complete solution then is

$$i = \frac{E}{Z}\sin(\omega t + \gamma - \theta) + A_1\epsilon^{a_1 t} + A_2\epsilon^{a_2 t} \tag{43}$$

in which A_1 and A_2 are integration constants and a_1 and a_2 are given in Eq. (28).

The capacitor voltage is

$$e_C = \frac{1}{C}\int i\,dt = -\frac{E}{\omega CZ}\cos(\omega t + \gamma - \theta) + \frac{A_1}{Ca_1}\epsilon^{a_1 t} + \frac{A_2}{Ca_2}\epsilon^{a_2 t} \tag{44}$$

Now applying the initial conditions of Eq. (38) at $t = 0$, there results from Eqs. (43) and (44), respectively

$$i(0) = I = \frac{E}{Z}\sin(\gamma - \theta) + A_1 + A_2 \tag{45}$$

$$e_C(0) = V = \frac{-E}{\omega CZ}\cos(\gamma - \theta) + \frac{A_1}{Ca_1} + \frac{A_2}{Ca_2} \tag{46}$$

Solving these two equations simultaneously for A_1 and A_2 and making use of Eq. (28), there results

$$A_1 = \frac{1}{\sqrt{R^2C^2 - 4LC}}\left\{\frac{I}{a_2} - CV - \frac{E}{Z}\left[\frac{1}{\omega}\cos(\gamma - \theta) + \frac{1}{a_2}\sin(\gamma - \theta)\right]\right\} \tag{47}$$

$$A_2 = \frac{-1}{\sqrt{R^2C^2 - 4LC}}\left\{\frac{I}{a_1} - CV - \frac{E}{Z}\left[\frac{1}{\omega}\cos(\gamma - \theta) + \frac{1}{a_1}\sin(\gamma - \theta)\right]\right\} \tag{48}$$

$$a_1 = \frac{-R}{2L} + \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}} \qquad a_2 = \frac{-R}{2L} - \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}} \tag{49}$$

These values, together with Eqs (41) and (42), inserted in Eqs. (43) and (44) give the solutions in a form suitable for the critically damped case However, if $R^2C < 4L$, then the radicals in Eqs. (47) to (49) become imaginary, and these expressions become complex numbers

$$A_1 = \tfrac{1}{2}(M + jN) \qquad A_2 = \tfrac{1}{2}(M - jN)$$
$$a_1 = -\alpha + j\omega \qquad a_2 = -\alpha - j\omega \tag{50}$$

and Eq (43) takes the damped oscillatory form

$$i = \frac{E}{Z}\sin(\omega t + \gamma - \theta) + \epsilon^{-\alpha t}(M\cos\omega t - N\sin\omega t) \tag{51}$$

It is evident from Eq (43) or Eq (51) that the final steady-state value after the transient has died out, is the alternating current of Eq (40). It is also clear from Eqs (47) and (48) that the amplitudes of the transient terms depend upon the angle γ on the ac applied voltage wave at $t = 0$

Resistance and inductance in series. This may be regarded as a special case of Eq (43) in which $C = \infty$ (the capacitance short-circuited) and $V = 0$ Under these conditions the solution becomes

$$i = \frac{E}{Z}\sin(\omega t + \gamma - \theta) + \left[I - \frac{E}{Z}\sin(\gamma - \theta)\right]\epsilon^{-(R/L)t} \tag{52}$$

Thus the transient starting from an initial value I decays exponentially to its final ac steady-state value There will be no transient if the switch is closed at an angle γ on the voltage wave such that

$$I = \frac{E}{Z}\sin(\gamma - \theta) \tag{53}$$

Resistance and capacitance in series. This may be regarded as a special case of Eq (43) in which $L = 0$ and $I = 0$. Under these conditions the solution becomes

$$i = \frac{E}{Z}\sin(\omega t + \gamma - \theta) - \left[\frac{V}{R} + \frac{E}{\omega CRZ}\cos(\gamma - \theta)\right]\epsilon^{-t/RC} \tag{54}$$

For a fixed value of emitter current I_e there is a fixed value of collector current αI_e added to the collector barrier leakage current I_{co}, giving a total collector current, $I_c = I_{co} + \alpha I_e$. This means that the slope of the dc characteristics should be the same as the slope of the collector-barrier leakage current curve for $I_e = 0$. The typical characteristics shown illustrate this. The slope of the collector leakage curve is very low since the collector voltage does not influence the relatively fixed number of minority carriers carrying the current. For a discussion of transistor characteristics, see TRANSISTOR CONNECTION.

High-frequency effects. These originate in three distinct properties of transistors: the transit time of injected carriers across the base region, the charging of the collector- or emitter-barrier capacitance through the base-region and collector-region resistances in series, and the time required to build up the proper density of injected carriers in the base region (called storage-capacity effect). In alloy junction transistors with a base region of uniform resistivity, the transport of injected carriers across the base is usually the limiting factor. Of course, base transit time alone introduces only a phase shift between the emitter and collector signals, but this time also gives a chance for injected carriers, bunched by the emitter signal, to diffuse apart and thus degrade the signal.

In drift transistors the base transit time is usually negligible compared to the charging time of the collector or emitter capacitance, and in some

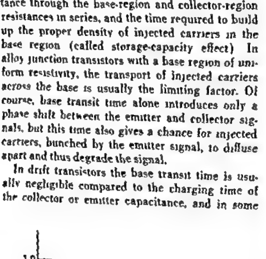


Fig. 3. Typical transistor frequency characteristics. (a) Frequency dependence of α. (b) Phase of collector current I_c versus emitter current I_e.

units the storage capacity seems to be an appreciable limitation.

Storage capacity also shows up in another way in transistors used as switches. Here it introduces a time delay both in turning on and in turning off the transistor. The turn-off delay is usually longer than the turn-on delay, because the density of injected carriers in the base region has had time to build up to large values during the time the transistor was on, and therefore takes a long time to subside to the level where the transistor can turn off. These delays are only slightly related to the actual time of rise or fall of the collector level, which is determined primarily by the collector-capacitance–base-resistance time constant.

Current multiplication. There are several transistor structures that give a collector-current multiplication factor α^* greater than unity. Two of these, the point-contact transistor and the *p-n-p-n* transistor, achieve the higher current multiplication using the same basic principle. This principle is called *p-n* hook multiplication.

The distinguishing feature of a hook structure is an electrically floating region of opposite conductivity type sandwiched bet…

base, the second *p*-region is floating; and the second *n*-region is the collector. In this structure the collector barrier is found between the base *n*-region and the floating *p*-region. This barrier, when reverse-biased, prevents the majority electrons of the base *n*-region from entering the floating *p* region but would not prevent any majority electrons of the collector *n*-region from entering the base if they could find their way across the floating *p*-region. When there is no emitter current, the barrier between the floating *p* region and the collector *n* region is forward-biased only enough to allow the collector barrier leakage current to flow. When emitter current flows, the injected holes traverse the base and enter the floating *p*-region. They charge up this region electrically so that the barrier between the floating p-region and the collector *n*-region becomes forward-biased and a large number of electrons flows across the floating *p*-region into the base. The current multiplication of this structure for low emitter currents depends upon the resistivities and effective thicknesses of the floating *p*-region and the collector *n*-region and can easily reach a value of $\alpha^* = 100$.

The multiplication factor can be designed into the *p-n-p-n* junction structure, but in the point contact transistor it depends on the efficacy of the electrical forming process. It may even be somewhat conjectural that the hook mechanism is responsible for the α^* of point-contact …

… for achieving an α^* greater than unity is to create extra current carriers in the region of the collector barrier by impact ionization. For example, in the case of a sim-

as germanium classifies the transistor by the fourth criterion. This type of transistor can be made in power ratings from 50 milliwatts (mw) to 30 watts, and in frequency ranges up to about 10 megacycles (Mc).

Another common transistor is the germanium *p-n-p* diffused-base drift transistor. Here the term *diffused-base refers to a fabrication technique* in which a resistivity gradient is produced in the base region by the use of solid-state diffusion of suitable impurities (*see* DIFFUSION IN SOLIDS). The term *drift refers to the principle of operation (third criterion)*, in which charge transport across the base region is accomplished by applying a built-in electric field to cause the charge to drift across, rather than to diffuse across, as in most other transistors. Similar terms are post alloy-diffusion drift transistor, pushed-out-base drift transistor, and graded-base transistor. This type of transistor normally has a power rating of about 50 mw and frequency ranges up to 700 Mc.

For further classification of transistors, *see* JUNCTION TRANSISTOR; POINT-CONTACT TRANSISTOR; SURFACE-BARRIER TRANSISTOR; see also the sections of this article on special devices and transistor manufacture.

Transistor action. To explain transistor action in more detail, some of the basic properties of a semiconductor material are first presented. An *n*-type semiconductor contains electrons and a

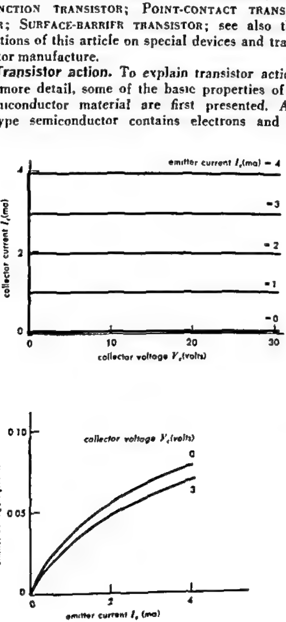


Fig 2. Typical transistor dc characteristics.

p-type semiconductor contains holes. These are called the majority carriers of the two types Actually there are always present a small number of holes in an *n*-type semiconductor and a small number of electrons in a *p*-type semiconductor. These are called the minority carriers of the two types. At a given temperature with a given material the product of the densities of the majority and minority carriers is a constant. This means that if there is present a very high density of majority carriers (low-resistivity material) there will be a correspondingly low density of minority carriers

The emitter current controls the collector current in a simple transistor. To understand this, first consider the magnitude of the collector current *in the absence of emitter current. In normal* operation the collector barrier is biased in the high-resistance (reverse) direction. Under this condition of bias the majority carriers are stopped by the barrier, and *only the minority carriers are free* to flow. If the collector barrier is a *p-n* junction and is composed of relatively low-resistivity material on both sides, this minority-carrier current can be of the order of microamperes or less. The emitter controls the collector current by varying the density of minority carriers on the base side of the *collector barrier.*

Injection. The emitter controls the density of minority carriers by injecting extra minority carriers into the base region when the emitter is biased in the low-resistance (forward) direction This is the fundamental process of simple transistor action Whenever a rectifying barrier is forward biased, extra minority carriers are added to the semiconductor near the barrier Since the source of these minority carriers is the majority-carrier density on the other *side of the barrier, it is clear that the* largest part of the forward current will be carried by those carriers which come from the largest-majority density A *p-n* junction will have a high injection efficiency for holes if the *p*-region has a much larger density of carriers (lower resistivity) than the *n*-region. Therefore, in a *p-n-p* transistor the emitter *p*-region should have a low resistivity compared to the *n*-type base region The phenomenon *of minority-carrier injection* is observed also in rectifying metal semiconductor contacts, and such contacts may be used as emitters as well as *p n* junctions.

Current gain The current gain α *of a simple* transistor may be expressed as the product of three factors: the fraction γ of the emitter current carried by the injected carriers, the fraction β of the injected carriers which arrive at the collector barrier, and the current multiplication factor α^* of the collector. For an alloy-junction transistor typical values of these factors are $\gamma = 0.980$, $\beta = 0.998$, and $\alpha^* = 1.000$, giving $\alpha = 0.978$ From *this it can be seen that most of the current which* flows into the emitter flows right on through the base region and out the collector, while only a small fraction (here 0.022) flows out the base connection.

lustrated I_c versus V_c characteristics gives a commonly used conductance h_{ce} which is the reciprocal of the output resistance with open-circuited (I_b = const) input.

The current gain h_{cb} (equal to $\partial I_c/\partial I_b$, with V_c = const) is related to current gain α (see TRANSISTOR) by $h_{cb} = \alpha/(1-\alpha)$. The input resistance h_{bb}, with a short-circuited (V_c = const) output, is given by the slope of the V_b versus I_b characteristics. Finally the voltage gain $1/h_{bc}$ (equal to $\partial V_c/\partial V_b$, with I_b = const) can be obtained from the separation of these curves. Typical values of these small signal parameters are h_{bb} = 1800 ohms; $1/h_{bc}$ = 1600; h_{cb} = 50, $1/h_{ce} = 0.05 \times 10^6$ ohms.

Common-base connection. This has an emitter-to-base input and a collector-to-base output connection. As with the common-emitter connection, the output resistance $1/h_{cc}$ is found from the slope h_{cc} of the I_c versus V_c characteristics. The current gain h_{ce} equals $\partial I_c/\partial I_e$ ($h_{ce} = -\alpha$ because of the assigned polarity of the currents; see Fig 2). The input resistance h_{ee} is found from the slope of the V_e versus I_e characteristics. The voltage gain $1/h_{ec}$ equals $\partial V_c/\partial V_e$. Typical values of the common-base parameters are h_{ee} = 36 ohms; $1/h_{ec}$ = 1500, $h_{ce} = -0.98$; $1/h_{cc} = 2.5 \times 10^6$ ohms.

Common-collector connection. This has a base-to-collector input and an emitter-to-collector output connection. Again the output resistance $1/h_{ee}$ is found from the slope h_{ee} of the I_e versus V_e characteristics. The current gain h_{eb} equals $\partial I_e/\partial I_b$. The input resistance h_{bb} is found from the slope of the V_b versus I_b characteristics. The voltage gain $1/h_{be}$ equals $\partial V_e/\partial V_b$. Typical values of the common-collector parameters are h_{bb} = 1800 ohms; $1/h_{be}$ = 1, $h_{eb} = -50$, $1/h_{ee} = 0.05 \times 10^6$ ohms.

Selection of transistor connections. From the foregoing definitions, it is possible to compare these connections for use in amplifiers. The power gain of the common-emitter connection (current gain times voltage gain) is seen to be the highest. This is the reason that this connection is used most frequently. The common-base connection shows the lowest input resistance and the highest output resistance. Its most useful characteristic is its linearity, which gives low distortion when driven by a

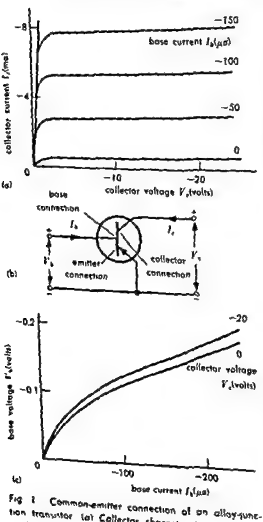


Fig 1 Common-emitter connection of an alloy-junction transistor (a) Collector characteristics (b) Schematic diagram (c) Base characteristics.

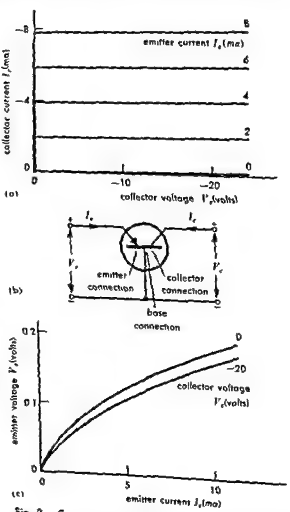


Fig 2 Common-base connection of an alloy-junction transistor (a) Collector characteristics. (b) Schematic diagram. (c) Emitter characteristics

ple transistor of the p-n-p type it was noted that almost all of the holes injected by the emitter traversed the base region and entered the collector giving $\alpha^* = 1.00$. If now the collector barrier has an electric field which imparts enough energy to the injected holes to ionize by impact some of the atoms of the material, the extra electrons so produced may flow from the collector to the base and increase the collector current for a given emitter current, thereby raising α^* above 1.00. This is called avalanche multiplication, and is the chief cause of the electrical breakdown of p-n junctions.

Transistor noise. Noise is quite low if a low source impedance is used. With source impedances of about 1000 ohms, a good junction transistor will have a noise factor of about 4 decibels. The noise factor is independent of the connection but rises with source impedances above 10,000 ohms and with frequencies below 1000 cycles.

Temperature effects. These are most marked in connection with the collector-barrier leakage current with no emitter current flowing I_{co}. This current increases exponentially with temperature and leads to a phenomenon called thermal runaway. If a transistor is operated at a given ambient temperature and a given initial power dissipation, this power will soon raise the temperature of the collector barrier, which then draws more current and in turn increases the dissipation. The process is cumulative and precautions must be taken to stabilize against it. Current gain increases slightly with increasing temperature in most n-p-n transistors, but this is a small effect unless the current gain is unusually close to unity.

Field-effect transistor. Also called unipolar transistor, this device does not operate by the process of injection and therefore is not a transistor in the normal sense. This device consists typically of a channel of relatively high-resistivity n-type semiconductor material which is constricted in the middle by a surrounding ring of low-resistivity p-type material. The ends of the channel carry ohmic contacts and the ring of p-type material (called the gate) carries a single ohmic contact A current is set up between the ends of the channel by external means and the gate is reverse biased relative to the input (source) end of the channel It is a property of a reverse-biased p-n junction between low- and high resistivity material, that the barrier region extends itself into the high-resistivity material as the voltage is increased. In this application an increased voltage on the gate will constrict the channel more and more until, at certain value of voltage (called the pinch-off voltage) the current through the channel is cut off. Variation of the gate voltage will modulate the channel current at voltages less than pinch-off. This device has a high input impedance compared to an ordinary transistor. Its characteristics resemble those of a vacuum tube pentode. Its frequency range is less than that of a good drift transistor.

Double-base diode. Also called a unijunction transistor, this consists of a single rectifying contact situated approximately midway along a semiconductor bar which carries two ohmic contacts at its ends. If a steady bias is applied between the ends of the bar a negative-resistance diode characteristic is observed between the rectifying contact and one end of the bar. This device is used primarily for switching.

Spacistor. This is a transistor in which the injection process takes place in the space-charge-free region of a reverse-biased p-n junction. Under these conditions, a high field exists at the point of injection so that the injected charges are swept away very rapidly. The device is difficult to stabilize and shows only moderate frequency response because of the relatively wide barrier region needed to accommodate the injecting contacts.

Transistor manufacture. The manufacture of transistors has required a whole new field of exacting technology. Good semiconductor material requires the maintenance of chemical purities far beyond the spectroscopic range A purity of one part in 10^9 is not unusual. Most devices must be made from oriented single crystals of semiconductor material which can have only very low densities of structural defects (*see* SINGLE CRYSTAL). Physical tolerances of the high-frequency transistor structures are microscopic; the separation of emitter and collector junctions must be of the order of a few microns in these units.

To solve these problems new techniques have appeared. Purity is achieved by melting a small zone of a bar, or ingot, and gradually passing this molten zone from one end of the bar to the other Impurities in the material remain in the liquid phase and are carried along with the molten zone, leaving high-purity material behind (*see* ZONE REFINING). Tolerances are achieved by a collection of new techniques, such as jet etching and plating (*see* SURFACE-BARRIER TRANSISTOR) solid-state diffusion (*see* JUNCTION TRANSISTOR), and the previously mentioned alloy-junction process [L.P.HU.]

Bibliography. L. P Hunter (ed.), *Handbook of Semiconductor Electronics*, 1956

Transistor connection

The method of connecting a transistor in a circuit The common-emitter, common base, and common-collector connections are the most frequently used and of these the common-emitter connection is by far the most popular for transistors with current gain $\alpha < 1.0$ To compare these connections, the following should be examined: small-signal current gain, voltage gain, input resistance with shorted output, and output resistance with open circuited input Shorted output means an ac short, the dc bias voltage is still present Open circuited input means a constant dc current bias These quantities are easily measured and are useful in calculating the performance of a transistor in a circuit connection

Common-emitter connection. This connection has a base-to-emitter input and a collector-to-emitter output connection The slope of the il

Transits of Mercury. These are relatively more frequent and occur at the rate of about 13 per century. The same geometric conditions are required as for transits of Venus, but the limits are not as narrow Conjunction must occur within three days of May 8 or within five days of November 10 Because of this, November transits are about twice as frequent as May transits. Observations of November transits yield more accurate results than those of May transits, because in November the motion of the planet is more rapid, thus permitting better timing of the contacts *See* MERCURY (PLANET)

Four contacts are observed: exterior ingress, interior ingress, interior egress, and exterior egress, designating respectively the exterior and interior points of tangency between the planet and the Sun at the beginning of the transit (ingress) and the interior and exterior points of tangency at the end (egress)

Transits of Mercury are observed for the purpose of determining the exact position of the planet and to improve data on the elements of its orbit.

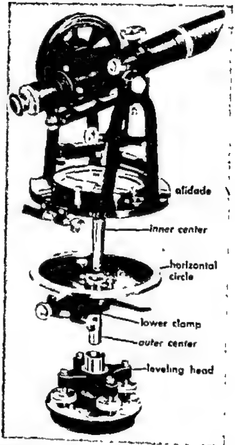
Transits of Jupiter's satellites. Transits of the Galilean satellites of Jupiter occur at each of their inferior conjunctions with the exception of satellite IV which occasionally passes clear of the planet's disk. They are difficult to observe and are used mainly to estimate the albedo (reflectivity) of the satellites relative to that of Jupiter. As each satellite passes in front of the planet, it casts its shadow on the planet's disk and causes the phenomenon of shadow transit. *See* JUPITER

Transits of stars. Passages of stars across the local *meridian* are observed extensively for solving *problems* of fundamental positional astronomy, timekeeping, and *navigation* At the precise instant of transit of a star across the local meridian, the local sidereal time is exactly *equal to the star's* right ascension, and the latitude of the observer is equal to the *sum of the* star's declination and its zenith distance. Furthermore, the difference between the local sidereal time and the Greenwich sidereal time gives a measure of the longitude of the observer Any of those quantities may be treated as the unknown to be determined by observations of meridian *transits See* ASTRONOMICAL COORDINATE SYSTEMS, ASTRONOMICAL INSTRUMENTS, ECLIPSE, ASTRONOMICAL. [S.D C.]

Transit (engineering)

A surveying instrument for measuring horizontal and vertical angles; also used for prolonging a straight line, establishing a level line of sight and measuring distances by the stadia method (*see* STADIA)

As seen in the disassembled view, three main elements support the telescope: at bottom, the leveling head, for making the vertical axis vertical; middle, the graduated horizontal circle (lower motion); top, the alidade, or upper motion, with verniers, level bubbles, and brackets for support of graduated vertical circle and telescope trunnions. *See* SURVEYING. [R H DO]



Engineer's transit disassembled. (*Keuffel and Esser Co*)

Transition elements

The members of the three series of chemical elements of atomic numbers 22–31, 40–49, and 72–81, inclusive, as they appear in the long or Bohr form of the *periodic* table. The name arises from their serving as *transitional* linkages between the most electropositive elements at one end of the series and the *least* electropositive at the other (*see* PERIODIC TABLE). In this sense, the transition elements include those specified above, but some authorities prefer to limit the serie...

there

bitals

Wh...

d...

21

in ... of electrons between *d* and *s* orbitals) at nickel, palladium, and platinum (Ni, Pd, and Pt, at. nos 28, 46, and 78)

The metals and their uses. All the transition elements are metals and, in general are ...

ized ...

vapo...

sten ... at 5510 and 3030°C. respectively At room temperature, the vapor pressure of tungsten is so low that it compares to only one gaseous atom in a volume of space equal to that of the known sidereal universe In general, those properties related to strong cohesiveness or

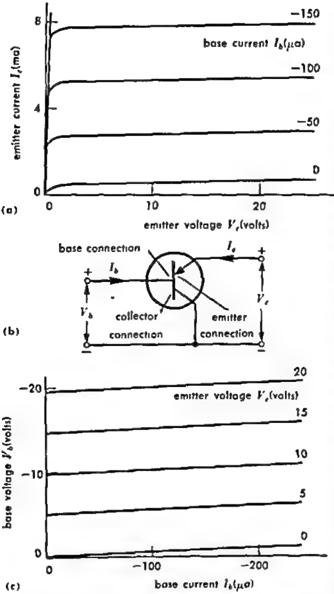


Fig. 3 Common-collector connection of an alloy-junction transistor. (a) Emitter characteristics. (b) Schematic diagram (c) Base characteristics.

current source. Accordingly, it is often used with a driver transformer. The common-collector connection shows the highest input resistance and the lowest output resistance. It is somewhat analogous to a cathode follower.

Switching circuits or pulse circuits can be either monostable, bistable, or astable. The input characteristic must have a negative-resistance region in order to be useful in such circuits. Such a characteristic may be achieved by a single transistor if it has a current gain $\alpha > 1.0$. Point-contact transistors have such an α and show a negative-resistance region in the base I_b versus V_b characteristic directly, or in the emitter I_e versus V_e characteristic if there is a sufficiently high resistance in the base circuit. The operation of the transistor will be monostable if the load line intersects this characteristic in only one point on one of the positive-resistance branches of the curve, bistable if the load line intersects the curve in two such points, and astable if it intersects the curve in only one point and that in the negative-resistance region The bistable circuit finds wide application in counters and computers. If transistors are used which have $\alpha < 1.0$, two transistors are required for each bistable switching circuit.

Complementary symmetry is the use of both *p-n-p* and *n-p-n* transistors together to take full advantage of their opposite bias and signal polarities For example, an emitter-follower circuit with base input and grounded collector will provide a positive drive to a load for a negative base signal in the case of a *p-n-p* transistor and for a positive base signal in the case of an *n-p-n* transistor. If the two are connected in parallel they will give a positive drive to a load with either polarity of input signal. [L.P.HU.]

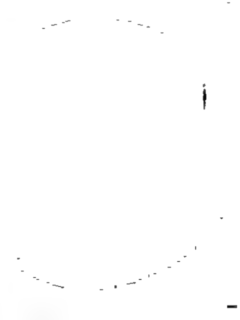
Bibliography: L. P. Hunter (ed.), *Handbook of Semiconductor Electronics*, 1956; R. F. Shea, *Principles of Transistor Circuits*, 1953.

Transit (astronomy)

The apparent passage of a planet across the surface of the Sun, of a satellite across the surface of the parent planet, or of a star, planet, or reference point across an adopted line of reference

Only Mercury and Venus are seen in transit across the surface of the Sun, because they are the only planets orbiting inside the path of Earth.

Transits of Venus. These are very rare. The Earth must be essentially in a straight line with Venus and the Sun, therefore Venus must be in inferior conjunction at the same time that it passes one of the nodes of its orbit. Conjunction must occur within two days of June 7 and December 9 to fulfill the conditions The last transit of Venus took place in 1882 There will not be another one until 2004 *See* VENUS; *see also* PLANET



Photograph of transit of Mercury on November 14, 1953 taken at Washington, D.C. (Official U S Navy Photograph)

Inasmuch as the *term* translucent seems to imply seeing, usage of the term is ordinarily limited to the visible region of the spectrum. *See* TRANSPARENT MEDIUM [M.G.M.]

Transmission, automotive

The device for providing different gear or drive ratios between the engine and drive wheels of an automotive vehicle. A principal function of the transmission is to enable the vehicle to accelerate up to a maximum through a wide speed range while the engine operates within its most effective speed range. The transmission is placed in the vehicle power train between the engine and the propeller shaft. Power passes through the transmission and propeller shaft and is delivered to the differential and drive axles. *See* AXLE; DIFFERENTIAL.

Transmission types. There are two general types of transmission, manual shift and automatic. Manual-shift transmissions employ layshaft or countershaft gearing and are used in conjunction with a clutch which disengages the engine from the transmission for shifting of gears in the transmission. *See* CLUTCH.

Automatic transmissions are of three general types: (1) fluid coupling with *planetary gearing* and automatic shifts, (2) torque converter with planetary gearing and automatic shifts, and (3) torque converter with planetary gearing without automatic shifts (*see* FLUID COUPLING; PLANETARY GEAR

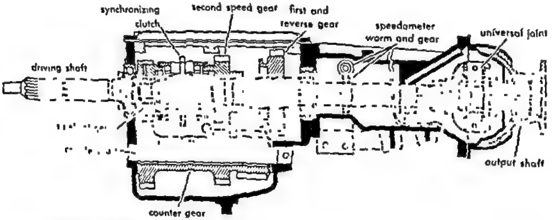


Fig. 1 Manual-shift transmission (Buick)

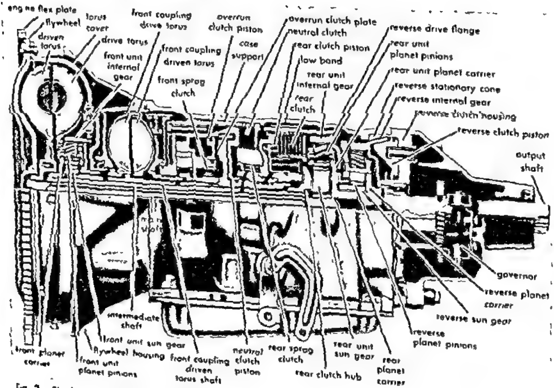


Fig. 2 Fluid-coupling automatic transmission (General Motors)

binding between the atoms in the metallic state, such as high density, extreme hardness, and high melting point, reach a broad maximum in the neighborhood of the third member of each series. Within a given subgroup, these same properties tend to increase with increasing atomic weight. Facility in the formation of metallic bonds is demonstrated also by the existence of a wide variety of alloys between different transition metals.

The transition elements include most metals of major economic importance, such as the relatively abundant iron, copper, nickel, and zinc, on one hand, and the rarer coinage metals, copper, silver, and gold, on the other. Also included are the rare and relatively unfamiliar element, rhenium, and technetium, which is not found naturally in the terrestrial environment, but is available in small amounts as a product of nuclear fission.

Chemical properties. In their compounds, the transition elements tend to exhibit multiple valency, the maximum valence increasing from +3 at the beginning of a series (Sc, Y, Lu) to +8 at the fifth member (Mn, Re). The higher valencies are attained more readily by the heavier elements; thereafter, the valencies decrease and do not exceed +3 for the common compounds of any of the three terminal members of each series (Cu, Ag, Au).

One of the most characteristic features of the transition elements is the ease with which most of them form stable complex ions. Features which contribute to this ability are favorably high charge- [illegible] plex of aurous gold $Au(Cn)_2^-$, of commercial importance for the recovery of the metal from low-grade ores, a similar complex $Ag(CN)_2^-$, useful in obtaining bright, firmly adherent deposits of silver by electroplating, and numerous ammonia complexes, of which the deep blue $Cu(NH_3)_4^{++}$ is a representative and familiar example, widely used in colorimetric analyses for copper. Vitamin B_{12} is an example of a cobalt(III) complex that is important in nutrition, and hemin, the red pigment of blood, is an important iron(II) complex.

Most of the ions and compounds of the transition metals are colored, and many of them are paramagnetic, that is, when placed in a magnetic field, the magnetic flux within the compound is higher than that of the surrounding field. Both color and paramagnetism are related to the presence of unpaired electrons in the *d* subshell. Excitation of these relatively loosely bound electrons to higher energy states accounts for the absorption of light in the visible region of the spectrum, while the magnetic field associated with the electron spin is responsible for the magnetic behavior of the compounds. Study of the magnetic behavior of compounds and complex ions of the transition elements has contributed much to an understanding of chemical bonding in these elements, since utilization of the *d* electrons in bonding involves electron-pair formation, with consequent cancellation of the magnetic moments and altered magnetic properties. Because of their ability to accept electrons in unoccupied *d* orbitals, transition elements and their compounds frequently exhibit catalytic properties. Many of the most important catalysts, such as nickel used in hydrogenation, are transition elements.

Broadly speaking, the properties of the transition elements are intermediate between those of the so-called representative elements, in which the subshells are completely occupied by electrons (alkali metals, halogen elements), and those of the inner or *f* transition elements, in which the subshell orbitals play a much less significant role in influencing chemical properties (rare-earth elements; actinide elements). *See* ATOMIC STRUCTURE AND SPECTRA; CATALYSIS; COMPLEX COMPOUNDS; COORDINATION CHEMISTRY; MAGNETOCHEMISTRY.

[B.B.CU.]

Bibliography: W. M. Latimer and J. H. Hildebrand, *Principles of Chemistry and Reference Book of Inorganic Chemistry*, 1941; T. Moeller, *Inorganic Chemistry*, 1952; L. Pauling, *College Chemistry*, 2d ed., 1955.

Transition point

The temperature at which a substance changes from one state of aggregation to another. This general definition would include the melting point (transition from solid to liquid), boiling point (liquid to gas), or sublimation point (solid to gas); but in practice the name transition point is usually restricted to the transition from one solid phase to another, that is, to the temperature at which a substance changes from one crystal structure to another.

Some typical examples of transition points are

β Fe (body-centered cubic)	→	γ-Fe (face-centered cubic) at 1180°K
S_8 (rhombic)	→	S_8 (monoclinic) at 369°K
CCl_4 (monoclinic)	→	CCl_4 (tetragonal) at 225.5°K
NH_4NO_3 (β rhombic)	→	NH_4NO_3 (α-rhombic) at 305.3°K
NH_4NO_3 (α rhombic)	→	NH_4NO_3 (trigonal) at 357.4°K

Another kind of transition point is the culmination of a gradual change (for example, the loss of ferromagnetism in iron or nickel) at the lambda point, or Curie point. This behavior is typical of second-order transitions. *See* BOILING POINT; EQUILIBRIUM, PHASE; FIRST-ORDER TRANSITION; MELTING POINT; SECOND-ORDER TRANSITION; SUBLIMATION; TRIPLE POINT.

[R.E.S.]

Translucent medium

A medium which transmits rays of light so diffused that objects cannot be seen distinctly, that is, the medium is only partially transparent. Familiar examples are various forms of glass which admit considerable light but impede vision.

Inasmuch as the term translucent seems to imply seeing, usage of the term is ordinarily limited to the visible region of the spectrum. See TRANSPARENT MEDIUM. [M.C.M.]

Transmission, automotive

The device for providing different gear or drive ratios between the engine and drive wheels of an automotive vehicle. A principal function of the transmission is to enable the vehicle to accelerate up to a maximum through a wide speed range while the engine operates within its most effective speed range. The transmission is placed in the vehicle power train between the engine and the propeller shaft. Power passes through the transmission and propeller shaft and is delivered to the differential and drive axles. See AXLE; DIFFERENTIAL.

Transmission types. There are two general types of transmission, manual shift and automatic. Manual shift transmissions employ layshaft or countershaft gearing and are used in conjunction with a clutch which disengages the engine from the transmission for shifting of gears in the transmission. See CLUTCH.

Automatic transmissions are of three general types: (1) fluid coupling with planetary gearing and automatic shifts, (2) torque converter with planetary gearing and automatic shifts, and (3) torque converter with planetary gearing without automatic shifts (see FLUID COUPLING; PLANETARY GEAR

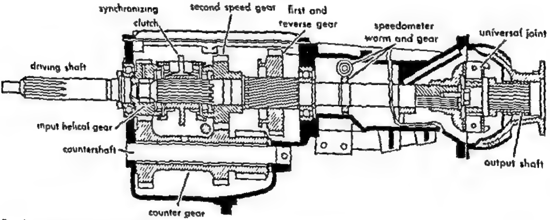


Fig. 1 Manual-shift transmission (Buick)

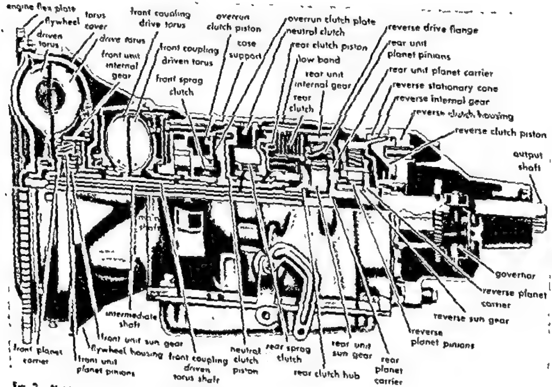


Fig. 2 Fluid-coupling automatic transmission (General Motors)

binding between the atoms in the metallic state, such as high density, extreme hardness, and high melting point, reach a broad maximum in the neighborhood of the third member of each series. Within a given subgroup, these same properties tend to increase with increasing atomic weight. Facility in the formation of metallic bonds is demonstrated also by the existence of a wide variety of alloys between different transition metals.

The transition elements include most metals of major economic importance, such as the relatively abundant iron, copper, nickel, and zinc, on one hand, and the rarer coinage metals, copper, silver, and gold, on the other. Also included are the rare and relatively unfamiliar element, rhenium, and technetium, which is not found naturally in the terrestrial environment, but is available in small amounts as a product of nuclear fission.

Chemical properties. In their compounds, the transition elements tend to exhibit multiple valency, the maximum valence increasing from +3 at the beginning of a series (Sc, Y, Lu) to +8 at the fifth member (Mn, Re). The higher valencies are attained more readily by the heavier elements; thereafter, the valencies decrease and do not exceed +3 for the common compounds of any of the three terminal members of each series (Cu, Ag, Au).

One of the most characteristic features of the transition elements is the ease with which most of them form stable complex ions. Features which contribute to this ability are favorably high charge-to-radius ratios and the availability of unfilled *d* orbitals which may be used in bonding. Examples of such complexes include a very stable cyanide complex of aurous gold $Au(Cn)_2^-$, of commercial importance for the recovery of the metal from low-grade ores, a similar complex $Ag(CN)_2^-$, useful in obtaining bright, firmly adherent deposits of silver by electroplating, and numerous ammonia complexes, of which the deep blue $Cu(NH_3)_4^{++}$ is a representative and familiar example, widely used in colorimetric analyses for copper. Vitamin B_{12} is an example of a cobalt(III) complex that is important in nutrition, and hemin, the red pigment of blood, is an important iron(II) complex.

Most of the ions and compounds of the transition metals are colored, and many of them are paramagnetic, that is, when placed in a magnetic field, the magnetic flux within the compound is higher than that of the surrounding field. Both color and paramagnetism are related to the presence of unpaired electrons in the *d* subshell. Excitation of these relatively loosely bound electrons to higher energy states accounts for the absorption of light in the visible region of the spectrum, while the magnetic field associated with the electron spin is responsible for the magnetic behavior of the compounds. Study of the magnetic behavior of compounds and complex ions of the transition elements has contributed much to an understanding of chemical bonding in these elements, since utilization of the *d* electrons in bonding involves electron-pair formation, with consequent cancellation of the magnetic moments and altered magnetic properties. Because of their ability to accept electrons in unoccupied *d* orbitals, transition elements and their compounds frequently exhibit catalytic properties. Many of the most important catalysts, such as nickel used in hydrogenation, are transition elements.

Broadly speaking, the properties of the transition elements are intermediate between those of the so-called representative elements, in which the subshells are completely occupied by electrons (alkali metals, halogen elements), and those of the inner or *f* transition elements, in which the subshell orbitals play a much less significant role in influencing chemical properties (rare-earth elements; actinide elements). *See* ATOMIC STRUCTURE AND SPECTRA; CATALYSIS; COMPLEX COMPOUNDS; COORDINATION CHEMISTRY; MAGNETOCHEMISTRY. [B.B CU.]

Bibliography: W. M. Latimer and J. H. Hildebrand, *Principles of Chemistry and Reference Book of Inorganic Chemistry*, 1941, T. Moeller, *Inorganic Chemistry*, 1952, L. Pauling, *College Chemistry*, 2d ed., 1955.

Transition point

The temperature at which a substance changes from one state of aggregation to another. This general definition would include the melting point (transition from solid to liquid), boiling point (liquid to gas), or sublimation point (solid to gas); but in practice the name transition point is usually restricted to the transition from one solid phase to another, that is, to the temperature at which a substance changes from one crystal structure to another.

Some typical examples of transition points are

β-Fe (body-centered cubic)	→	γ-Fe (face-centered cubic) at 1180°K
S_8 (rhombic)	→	S_8 (monoclinic) at 369°K
CCl_4 (monoclinic)	→	CCl_4 (tetragonal) at 225.5°K
NH_4NO_3 (β-rhombic)	→	NH_4NO_3 (α-rhombic) at 305.3°K
NH_4NO_3 (α rhombic)	→	NH_4NO_3 (trigonal) at 357.1°K

Another kind of transition point is the culmination of a gradual change (for example, the loss of ferromagnetism in iron or nickel) at the lambda point, or Curie point. This behavior is typical of second-order transitions. *See* BOILING POINT; EQUILIBRIUM, PHASE; FIRST-ORDER TRANSITION; MELTING POINT; SECOND-ORDER TRANSITION; SUBLIMATION; TRIPLE POINT. [R.L.S.]

Translucent medium

A medium which transmits rays of light so diffused that objects cannot be seen distinctly, that is, the medium is only partially transparent. Familiar examples are various forms of glass which admit considerable light but impede vision.

The first movement of the sliding clutch when shifting into second or direct engages a bronze cone with a mating steel cone on the second-speed or the direct gear. This contact turns the cone assembly, causing the chamfers on the three pins to engage with the chamfers in the three holes in the sliding clutch.

Further pressure on the cones as the operator continues the shifting action causes the second or direct-drive gear speed to come to the same speed as the sliding clutch or output shaft speed, thus synchronizing with it. This permits the chamfers on the clutch to slide along the chamfers on the pins, thereby relieving pressure on the cones and permitting the clutch teeth to engage with the corresponding teeth on the second speed or direct-drive gears.

Fluid-coupling transmission. The fluid-coupling transmission using planetary automatic shifting, called Hydramatic, was first used in the 1940 Oldsmobile (Fig. 2). Operation is shown in Fig. 3 by a series of line diagrams that indicate the flow of the torque or turning effect in the four speeds and reverse. The large fluid coupling is always filled. The small coupling is filled or emptied, depending on whether the drive is through it or not. The lines indicate the flow of torque through the transmission. Application of clutching and fluid coupling to obtain four speeds is shown in Fig. 3.

The shifts from one gear ratio to another are made by a hydraulic control system in accordance with engine throttle opening and car speed. On light throttle the shift is early; on full throttle it is late, as shown in Table 2. Throttle effect comes from accelerator position. Car-speed effect comes from two hydraulic centrifugal governors, which deliver pressure somewhat proportional to car speed.

Table 2 Speeds at which shifts occur for fluid-coupling

Shifts	Speed, mph	
	Minimum throttle	Maximum throttle
1 to 2	7	25
2 to 3	15	43
3 to 4	22	78

For coasting downhill at 2.55:1 ratio, a separate rear brake is provided. This brake locks the rear ring gear to the case to prevent the rear one-way clutch from overrunning when the rear wheels drive the engine. Likewise, a disk brake is provided for coasting downhill at a 1.55:1 ratio; the front one-way clutch would also overrun under this condition.

Torque-converter transmissions. A variety of torque-converter transmissions are used. However, they can be grouped under two basic types: (1) automatic shifting and (2) nonshifting (Table 1).

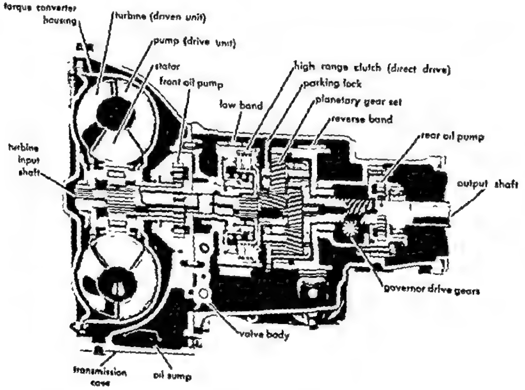


Fig. 4 Two-speed automatic torque-converter transmission (Chevrolet).

TRAIN; TORQUE CONVERTER). Table 1 shows transmissions used in American cars in 1959.

Table 1. Transmissions used on American cars

	Manual shift	Automatic shifts		Automatic (no shifts)
	Layshaft transmission	Fluid coupling	Torque converter	Torque converter
American Motors	×		×	
Buick	×			×
Cadillac		×		
Chevrolet	×		×	×
Chrysler	×		×	
DeSoto	×		×	
Dodge	×		×	
Edsel	×		×	
Ford	×		×	
Lincoln	×		×	
Mercury	×		×	
Oldsmobile	×	×		
Plymouth	×		×	
Pontiac	×	×		
Studebaker-Packard	×		×	
Willys	×			

In Great Britain and Europe transmissions are mainly the layshaft type of manual shift. The preselecting transmission employing a fluid coupling and planetary gearing has also been used by Daimler in England for many years. Rolls-Royce and Bentley in England employ a transmission of the fluid-coupling planetary-gear type with automatic shifts. Armstrong Sidley, British Motor Car Group (most models), Ford, Jaguar, Rootes Group (larger models), Rover, and Standard in England, and Daimler-Benz in Germany use a torque converter with planetary gearing and automatic shifts. Some makers employ automatic clutches with manual-shifting layshaft units.

Manual shift. Figure 1 is a section through a manual-shift transmission employing layshaft gearing typical of American passenger cars.

Shift mechanism. The transmission consists of a driving shaft at the left, connected to the engine, usually by a foot-operated single-plate clutch. The input shaft carries a helical gear meshing with the driven gear on the countershaft. The countershaft carries two more gears, as shown. The second-speed gear meshes with a gear on the output shaft. The first-speed gear meshes with the first and reverse sliding gear on the output shaft when the latter is engaged with it to obtain first speed. The first-speed gear meshes with an idler gear, which in turn meshes with the first and reverse sliding gear on the output shaft when the latter is engaged with it. The idler is used to reverse rotation of the output shaft.

First speed is thus obtained by sliding the gear on the extreme right of the output shaft to the left, giving a 2.5:1 speed ratio; reverse is obtained by sliding it to the right, giving a 3:1 ratio. The teeth of the meshing gears are rounded and chamfered at their engaging ends to permit easy engagements.

Second speed is obtained by sliding the synchronizing clutch located between the input gear and the second speed gear to the right to engage with clutch teeth integral with the second speed gear, giving a 1.5:1 speed ratio; direct drive (third speed) is obtained by sliding it to the left to engage with clutch teeth integral with the gear on the input shaft, giving a 1:1 speed ratio. The sliding clutch is movably splined to the output shaft to complete the drive train.

Synchronizing clutch. To avoid clashing when the sliding clutch engages the second-speed gear or the direct-drive gear, synchronizing is employed to bring the second-speed or direct-drive gears to the speed of the sliding clutch splined on the output shaft. The synchronizer consists of two bronze cones held together by three pins passing through three chamfered holes in a center flange integral with the sliding clutch. The pins have central grooves with chamfered sides. The synchronizer is centered by springs (assisted by the cones).

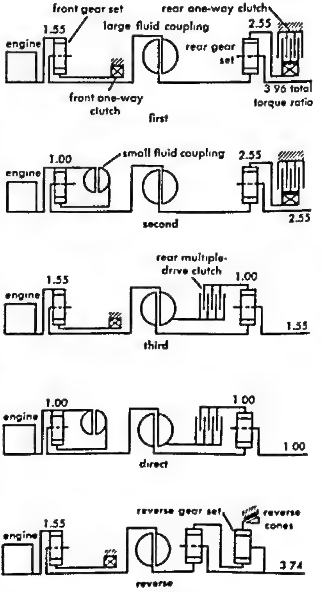


Fig. 3 Power flow through fluid-coupling automatic transmission.

The first movement of the sliding clutch when shifting into second or direct engages a bronze cone with a mating steel cone on the second-speed or the direct gear. This contact turns the cone assembly, causing the chamfers on the three pins to engage with the chamfers in the three holes in the sliding clutch

Further pressure on the cones as the operator continues the shifting action causes the second or direct-drive gear speed to come to the same speed as the sliding clutch or output shaft speed, thus synchronizing with it This permits the chamfers on the clutch to slide along the chamfers on the pins, thereby relieving pressure on the cones and permitting the clutch teeth to engage with the corresponding teeth on the second-speed or direct-drive gears

Fluid-coupling transmission. The fluid-coupling transmission using planetary automatic shifting, called Hydramatic, was first used in the 1940 Oldsmobile (Fig 2) Operation is shown in Fig 3 by a series of line diagrams that indicate the flow of the torque or turning effect in the four speeds and reverse The large fluid coupling is always filled. The small coupling is filled or emptied, depending on whether the drive is through it or not. The lines indicate the flow of torque through the transmission Application of clutching and fluid coupling to obtain four speeds is shown in Fig 3

The shifts from one gear ratio to another are made by a hydraulic control system in accordance with engine throttle opening and car speed. On light throttle the shift is early; on full throttle it is late, as shown in Table 2. Throttle effect comes from accelerator position Car-speed effect comes from two hydraulic centrifugal governors, which deliver pressure somewhat proportional to car speed

Table 2 Speeds at which shifts occur for fluid-coupling

Shifts	Speed, mph	
	Minimum throttle	Maximum throttle
1 to 2	7	25
2 to 3	15	43
3 to 4	22	78

For coasting downhill at 2.55:1 ratio, a separate rear brake is provided This brake locks the rear ring gear to the case to prevent the rear one-way clutch from overrunning when the rear wheels drive the engine. Likewise, a disk brake is provided for coasting downhill at a 1.55:1 ratio; the front one-way clutch would also overrun under this condition

Torque-converter transmissions. A variety of torque-converter transmissions are used. However, they can be grouped under two basic types: (1) automatic shifting and (2) nonshifting (Table 1).

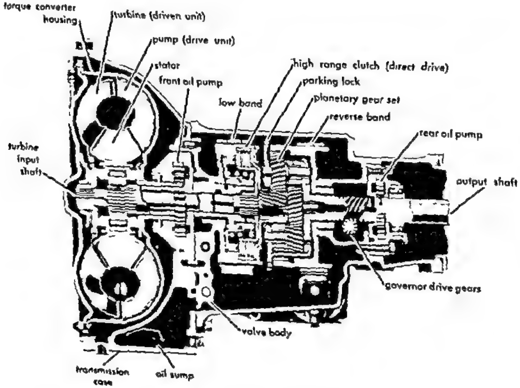


Fig 4 Two-speed automatic torque-converter transmission (Chevrolet)

TRAIN; TORQUE CONVERTER). Table 1 shows transmissions used in American cars in 1959.

Table 1. Transmissions used on American cars

	Manual shift	Automatic shifts		Automatic (no shifts)
	Layshaft transmission	Fluid coupling	Torque converter	Torque converter
American Motors	×		×	
Buick	×			×
Cadillac		×		
Chevrolet	×		×	×
Chrysler	×		×	
DeSoto	×		×	
Dodge	×		×	
Edsel	×		×	
Ford	×		×	
Lincoln	×		×	
Mercury	×		×	
Oldsmobile	×	×		
Plymouth	×		×	
Pontiac	×	×		
Studebaker-Packard	×		×	
Willys	×			

In Great Britain and Europe transmissions are mainly the layshaft type of manual shift The preselecting transmission employing a fluid coupling and planetary gearing has also been used by Daimler in England for many years. Rolls-Royce and Bentley in England employ a transmission of the fluid-coupling planetary-gear type with automatic shifts. Armstrong Sidley, British Motor Car Group (most models), Ford, Jaguar, Rootes Group (larger models), Rover, and Standard in England, and Daimler-Benz in Germany use a torque converter with planetary gearing and automatic shifts. Some makers employ automatic clutches with manual-shifting layshaft units

Manual shift. Figure 1 is a section through a manual-shift transmission employing layshaft gearing typical of American passenger cars.

Shift mechanism. The transmission consists of a driving shaft at the left, connected to the engine, usually by a foot-operated single-plate clutch The input shaft carries a helical gear meshing with the driven gear on the countershaft. The countershaft carries two more gears, as shown. The second-speed gear meshes with a gear on the output shaft. The first-speed gear meshes with the first and reverse sliding gear on the output shaft when the latter is engaged with it to obtain first speed. The first-speed gear meshes with an idler gear, which in turn meshes with the first and reverse sliding gear on the output shaft when the latter is engaged with it The idler is used to reverse rotation of the output shaft.

First speed is thus obtained by sliding the gear on the extreme right of the output shaft to the left, giving a 2.5:1 speed ratio; reverse is obtained by sliding it to the right, giving a 3:1 ratio. The teeth of the meshing gears are rounded and chamfered at their engaging ends to permit easy engagements.

Second speed is obtained by sliding the synchronizing clutch located between the input gear and the second speed gear to the right to engage with clutch teeth integral with the second speed gear, giving a 1.5:1 speed ratio; direct drive (third speed) is obtained by sliding it to the left to engage with clutch teeth integral with the gear on the input shaft, giving a 1:1 speed ratio. The sliding clutch is movably splined to the output shaft to complete the drive train.

Synchronizing clutch. To avoid clashing when the sliding clutch engages the second-speed gear or the direct-drive gear, synchronizing is employed to bring the second-speed or direct-drive gears to the speed of the sliding clutch splined on the output shaft. The synchronizer consists of two bronze cones held together by three pins passing through three chamfered holes in a center flange integral with the sliding clutch. The pins have central grooves with chamfered sides. The synchronizer is centered by springs (assisted by the cones).

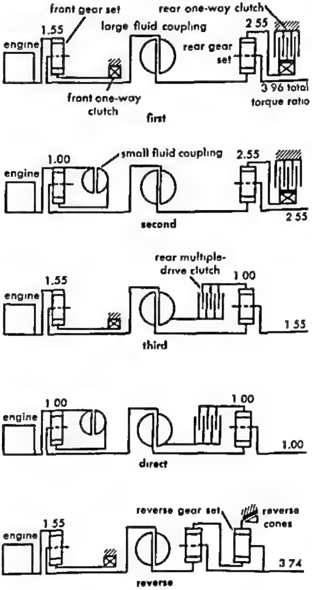


Fig 3 Power flow through fluid-coupling automatic transmission

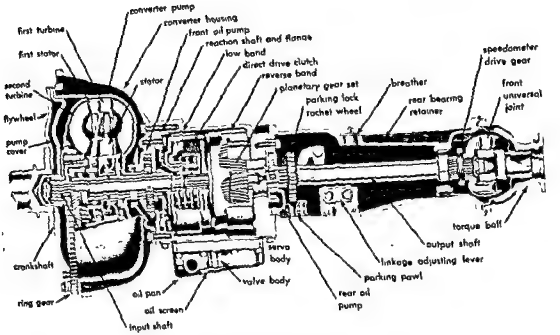


Fig 7 Twin-turbine nonshifting torque-converter transmission (Buick)

bine (geared 1:1 with the output shaft) gradually assumes its portion of the torque and finally takes over the drive completely (Fig 8).

When the driver depresses the accelerator pedal to the toeboard for increased acceleration, a hydraulic piston rotates the second reactor blades from low to high pitch, thereby increasing the torque output of the converter

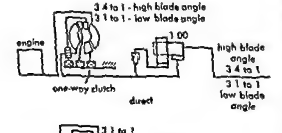


Fig 8 Power flow through nonshifting transmission

Forward, neutral, reverse, low, and park positions are manually selected In forward drive, the rear planetary gear set is locked up, giving a 1:1 drive ratio through it In reverse the ring gear is locked to the case. In low range (used for hill-braking) the front sun gear is locked to the case. Both the reverse and hill-braking positions have a ratio of 1 8 1. [F.R.M.]

Bibliography W. H. Crouse, *Automotive Transmissions and Power Trains*, 2d ed., 1959; K W. Gage and P J Rhoads, *SAE Trans*, vol. 65, 1957; R J Gorsky, Buick's twin turbine Dynaflow transmission, *SAE Trans.*, vol 63, 1955, R. S. Plexico and R E Kaufman, Chevrolet automatic transmission, *SAE Quart Trans*, vol 4, 1950; W. R. Rodgers and A J Syrovy, The Chrysler and Powerflite transmission *SAE Trans.*, vol 62, 1954; H T Youngren and H G English, Ford-Mercury automatic transmission, *SAE Quart Trans.*, vol 4, 1950

Transmission lines

A system of conductors suitable for conduction of electric power or signals between two or more termini For example, commercial-frequency electric power transmission lines connect electric generating plants substations, and their loads Telephone transmission lines interconnect *telephone* subscribers and telephone exchanges Radio-frequency transmission lines transmit high-frequency

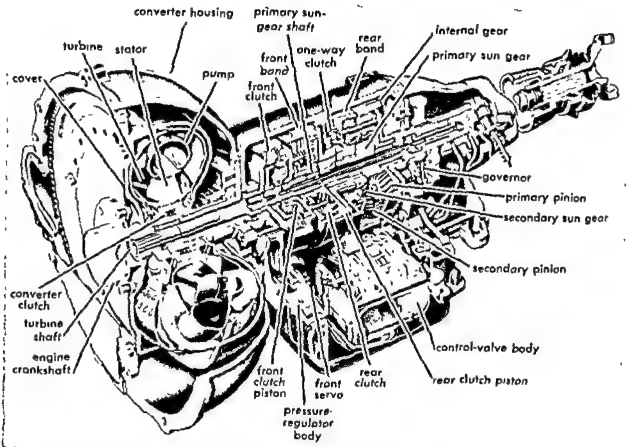


Fig 5. Three-speed automatic torque-converter transmission (Ford)

Automatic shifting. This type of transmission appears in two forms: (1) the two-speed unit, with ratios of approximately 1.8:1 and 1:1, and (2) the three-speed unit with ratios of approximately 2.5:1, 1.5:1, and 1:1 Representative models are shown in Figs. 4 and 5. Power flow through a representative three-speed transmission is shown in Fig 6 Shifting through the speeds is accomplished by engine-throttle, car-speed-sensitive means similar to the construction used in the automatic-shifting fluid-coupling transmission. Forward, reverse, hill-braking, and park positions are selected manually.

Hill-braking is accomplished by locking the transmission in the lower ratios, thereby avoiding freewheeling.

Converters in automatic-shifting transmissions are of the three-element type, consisting of pump, turbine, and reactor. Torque multiplication is 2:1 or greater at stall speed with full throttle (car stationary) and decreases as the car accelerates.

Nonshifting. Changes in speed ratios in non-shifting transmissions are accomplished by using a converter arrangement of pump, two driven turbines, and two reactors, the second of which has two blade positions (Fig 7).

The first turbine, adjacent to the driving pump, is fastened to the ring gear in the planetary gear set, in the converter housing, at a ratio of 1.6:1 to the output shaft. The second turbine is mounted to the planetary gear-set cage, which runs at output shaft speed.

As the car begins to accelerate, the first turbine furnishes the major portion of the driving torque. As the car continues to accelerate, the second tur-

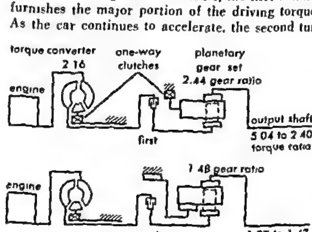


Fig 6. Power flow in two-speed automatic transmission

Principal mode. When the electric and magnetic field vectors are perpendicular to one another and *transverse* to the direction of the transmission line, this condition is called the principal mode or the transverse electromagnetic (TEM) mode. The principal mode electric- and magnetic-field configurations about the conductors are essentially those of Fig. 1. Modes other than the principal mode may exist at any frequency for which conductor spacing exceeds one half of the wavelength of an electromagnetic wave in the medium separating the conductors. Such high-frequency modes are called wave-guide transmission modes. See WAVE GUIDE.

In a uniform [illegible]

wave. In a line with negligible losses the transmitted shape remains unchanged. When losses are present, the shape, unless sinusoidal, is altered, because the phase velocity and attenuation vary with frequency. See PHASE VELOCITY.

If sinusoidal, the voltage and current decay exponentially as a wave progresses. The voltage or current, at a distance x from the sending end, is decreased in magnitude by a factor of $\varepsilon^{-\alpha x}$, where ε is the Napierian base and α is called the *attenuation constant*. The voltage or current at that point lags behind the voltage or current at the sending end by the phase angle βx, where β is called the phase constant.

The attenuation constant α and the phase constant β depend on the distributed parameters of the transmission line, which are (1) resistance per unit length r, the series resistance of a unit length of both going and return conductors, (2) conductance per unit length g, the leakage conductance of the insulators, conductance due to dielectric losses, or both, (3) inductance per unit length l, determined as flux linkages per unit length of a line of infinite extent carrying a constant direct current, and (4) capacitance per unit length c, determined from charge per unit length of a line of infinite extent with constant voltage applied.

The values of α and β may be found from the complex equation

$$\alpha + j\beta = \sqrt{(r + j2\pi fl)(g + j2\pi fc)}$$

where j is the notation for the imaginary number $\sqrt{-1}$ and f is the frequency of the alternating voltage and current. The complex quantity $\alpha + j\beta$ is often called the propagation constant γ. Since $r + j2\pi fl$ is the impedance z per unit length of line and $g + j2\pi fc$ is the admittance y per unit length of line, the equation for the propagation constant is often written simply as

$$\gamma = \sqrt{zy}$$

The velocity at which a point of constant phase is propagated is called the phase velocity v, and is equal to $2\pi f/\beta$. For negligible losses in the line (when r and g are approximately zero) the phase velocity is $1/\sqrt{lc}$, which is also the velocity of electromagnetic waves in the medium surrounding the transmission-line conductors.

The distributed inductance and resistance of the lines may be modified from their dc values because of skin effect in the conductors. This effect, which increases with frequency and conductor size, is usually, but not always, negligible at power frequencies. See SKIN EFFECT.

Characteristic impedance. The ratio of the voltage to the current in either the forward or the reflected wave is the complex quantity Z_0, called the characteristic impedance.

When line losses are relatively low, that is, when

$$r \ll 2\pi fl$$
$$g \ll 2\pi fc$$

the characteristic impedance becomes

$$Z_0 = \sqrt{l/c}$$

or a quantity nearly independent of frequency (but not exactly so since both l and c may be somewhat frequency-dependent). The magnitude of Z_0 is used widely, at high frequencies, to identify a type of transmission line such as 50-ohm line, 200-ohm line, and the popular 300-ohm antenna lead-in line used with tele[illegible]

for [illegible]

possible of signal waveshape distortion. Two types of distortion occur. One is a form of amplitude distortion due to line attenuation, which varies with the signal frequency. The other, delay distortion, occurs when the component frequencies of a signal arrive at the receiving end at different instants of time. This occurs because the velocity of propagation along the line is a function of the frequency.

Theoretically, a distortionless line can be devised if the line parameters are adjusted so that $r/g = l/c$. In practice this is approached by employing loading circuits. Under these conditions the propagation constant becomes

$$\gamma = \alpha + j\beta = \sqrt{r/g}(g + j2\pi fc)$$

The attenuation constant α is $\sqrt{rg}$, which is independent of frequency f. Therefore, there will be no frequency distortion.

The phase constant β is $2\pi f\sqrt{lc}$ which depends upon frequency. The velocity of propagation along any transmission line is $2\pi f/\beta$, and for the distortionless line this becomes $1/\sqrt{lc}$. Thus the velocity of propagation is independent of frequency, and there will be no delay distortion.

Transmission-line equations. The principal-mode properties of the *transmission-line* equations are described by the equations

$$\frac{\partial e}{\partial x} = -\left(ri + l\frac{\partial i}{\partial t}\right) \qquad (1)$$

$$\frac{\partial i}{\partial x} = -\left(ge + c\frac{\partial e}{\partial t}\right) \qquad (2)$$

electric signals between antennas and transmitters or receivers. The theory of transmission lines is considered first, followed by its application to power transmission lines. *See* TRANSMISSION THEORY AND METHODS.

Although only a short cord is needed to connect an electric lamp to a wall outlet, the cord is, properly speaking, a transmission line. However, in the electrical industry the term *transmission line* is applied only when both voltage and current at one line terminus may differ appreciably from those at another terminus as a result of the electrical properties of the line. Transmission lines are described as (1) electrically short, if the difference between terminal conditions are attributable simply to the effects of conductor series resistance and inductance, or to the effects of a shunt leakage resistance and capacitance, or to both results; and (2) electrically long, when the properties of the line result from traveling-wave phenomena.

TRANSMISSION-LINE THEORY

Depending on the configuration and number of conductors and the electric and magnetic fields about the conductors, transmission lines are described as open-wire transmission lines, coaxial transmission lines, cables, or wave-guide transmission lines.

Open-wire transmission lines. Open-wire lines may comprise a single wire with an earth (ground) return or two or more conductors. The conductors are supported at more or less evenly spaced points along the line by insulators, with the spacing between conductors maintained as nearly uniform as feasible, except in special-purpose tapered transmission lines, discussed later in this section.

Open-wire construction is used for communication or power transmission whenever practical and permitted, as in open country and where not prohibited by ordinances.

Open-wire lines are economical to construct and maintain and have relatively low losses at low and *medium frequencies*. Difficulties arise from electromagnetic radiation losses at very high frequencies and from inductive interference, or crosstalk, resulting from the electric and magnetic field coupling between adjacent lines accompanying the characteristic field configuration (Fig. 1).

Coaxial transmission lines. A coaxial transmission line comprises a conducting cylindrical shell, solid tape, or braided conductor surrounding an isolated, concentric, inner conductor which is solid, stranded, or (in certain video cables and delay cables) helically wound on a plastic or ferrite core. The inner conductor is supported by ceramic or plastic beads or washers in air- or gas-dielectric lines, or by a solid polyethylene or polystyrene dielectric.

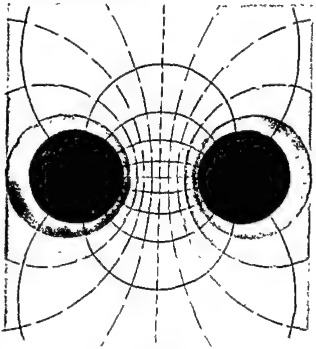
The purpose of this construction is to have the shell prevent radiation losses and interference from external sources. The electric and magnetic fields shown in Fig. 1*b* are nominally confined to the space inside the outer conductor. Some external fields exist, but may be reduced by a second outer sheath.

Coaxial lines are widely used in radio, radar, television, and similar applications.

Sheathed cables. Also termed shielded cables, these comprise two or more conductors surrounded by a conducting cylindrical sheath, commonly supported by a continuous solid dielectric. The sheath provides both shielding and mechanical protection.

Coaxial lines, sheathed cables, or shielded cables are often termed simply cables. For cable assemblies of coaxial lines and other circuits *see* COAXIAL CABLE.

Traveling waves. When electric power is applied at a terminus of a transmission line, electromagnetic waves are launched and guided along the line. The steady-state and transient electrical properties of transmission lines result from the superposition of such waves, termed direct waves, and the reflected waves which may appear at line discontinuities or at load terminals.


(a)

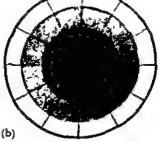

(b)

Fig. 1 Directions of the electric (solid lines) and magnetic (dashed lines) fields about two-conductor (a) open-wire and (b) coaxial transmission lines in a plane normal to the conductors, for continuous and low-frequency currents.

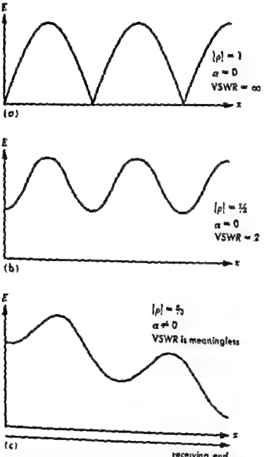


Fig 3 Voltage distribution under sinusoidal steady-state conditions on a section of transmission line, illustrating three standing-wave conditions: (a) line with negligible losses, reflection coefficient of unity, (b) line with negligible losses, reflection coefficient of 1/3, and (c) line with finite losses, reflection coefficient of 3/5. Position of the voltage wave in each case is dependent on the angle of the phasor value of reflection coefficient. In each case a current maximum (not shown) appears at a voltage minimum.

POWER TRANSMISSION LINES

In an electric power system the facility used to transfer large amounts of power from one location to a distant location is termed a power transmission line. Techniques of power transmission are presented in this section.

Power transmission lines nearly always operate with three-phase alternating currents having a frequency of 60 cycles per second (cps). This is almost universal in the United States. Three conductors usually comprise the line. For transmitting large amounts of power over long distances, high voltages are necessary. Standard transmission voltages are 11, 22, 33, 44, 66, 110, 132, 154, 220, 287, and 330 kilovolts (kv). These are nominal values; the actual voltages vary from sending end to receiving end and from time to time. The line conductors are usually placed overhead, supported by poles or towers; however, in special circumstances they form part of a cable under the ground or under water. *See* ELECTRIC POWER SYSTEMS.

Requirements of transmission. Power transmission systems must be reliable, have good voltage regulation and adequate power capability, and be capable of economical operation.

Reliability. This requirement is met by sturdy construction, by protection against overvoltages, by rapid automatic disconnection of accidentally short-circuited lines, by suitable transmission layouts, and by automatic rapid reconnection of lines experiencing only transitory faults.

Good voltage regulation. When the load voltage does not vary appreciably as the load increases from no load to full load, the regulation is said to be good. The inherent voltage regulation depends mainly on the inductive reactance of the line and the power factor of the load. If the inherent regulation is unsatisfactory, the voltage can be controlled by switched shunt capacitors or synchronous condensers connected at the [illegible]

[illegible]

posed by losses, temperature of the conductors, voltage regulation, and system stability, is the power capability of the line. It varies approximately as the square of the voltage and inversely as the distance. Typical capabilities of single-circuit 60 cycle aerial lines are shown in the table.

Power capability of 60-cycle aerial lines

Voltage, kilovolts	Length, miles	Capability, megawatts
11	10	1
66	60	5
132	120	60
220	240	150

Economy. Fulfillment of this requirement depends on a balance between low first cost and low operating cost, including cost of power loss. The principal loss is the I^2R loss in the conductors.

Constants. From a knowledge of the size and type of conductors and the spacing between them, one can obtain the values of series resistance r and inductive reactance x per phase per unit length of line and of shunt capacitive susceptance b and leakage conductance g per phase per unit length of line. All of these values are multiplied by the length of the line, giving constants R, X, B, and G respectively. These are then combined to give the complex impedance, admittance, and phase displacement, respectively

$$Z = R + jX \qquad Y = G + jB \qquad \theta = \sqrt{ZY}$$

For power lines it may be assumed that $G = 0$.

Equivalent circuit. This circuit indicates lumped values which represent values distributed along the line. A short line can be represented adequately by its nominal π circuit shown in Fig. 4a. Here the shunt admittance Y, actually distributed uniformly

in which e and i are instantaneous values of voltage and current respectively, x is distance from the sending terminals, and t is time

For steady-state sinusoidal conditions, the solutions of these equations are

$$E = E_s \cosh \gamma x - I_s Z_0 \sinh \gamma x$$
$$I = I_s \cosh \gamma x - \frac{E_s}{Z_0} \sinh \gamma x$$

for voltage E and current I at a distance x from the sending end in terms of voltage E_s and current I_s at the sending end.

In terms of receiving-end voltage E_r and current I_r, these solutions are

$$E = E_r \cosh \gamma x + I_r Z_0 \sinh \gamma x$$
$$I = I_r \cosh \gamma x + \frac{E_r}{Z_0} \sinh \gamma x$$

where x is now the distance from the receiving end.

Reflection coefficient. If the load at the receiving end has an impedance Z_r, the ratio of reflected voltage to direct voltage, known as the reflection coefficient ρ, is

$$\rho = \frac{Z_r - Z_0}{Z_r + Z_0}$$

When the load impedance is equal to Z_0 the reflection coefficient is zero. Under this condition the line is said to be matched. *See* REFLECTION AND TRANSMISSION COEFFICIENTS

Pulse transients. The transient solutions of Eqs. (1) and (2) are dependent on the particular problem involved. Typical physical phenomena with pulse transients are shown in Fig 2 The characteristic time delay in transmission is often advantageously employed in radar and other pulse-signal systems For examples, *see* DELAY LINE.

Standing waves. The superposition of direct and reflected waves under sinusoidal conditions in an [illegible] (Fig 3). [illegible] es are negligible, successive maxima are approximately equal, under this condition a quantity, voltage standing-wave ratio, abbreviated VSWR, is defined as

$$\text{VSWR} = \frac{V_{max}}{V_{min}}$$

Power standing-wave ratio. This quantity, abbreviated PSWR, is equal to $(\text{VSWR})^2$. Measurements of voltage magnitude and distribution on a line of known characteristic impedance Z_0 can be used to determine the magnitude and phase angle of an unknown impedance connected at its receiving end. Lines adapted for such impedance measurements, [illegible] mped- [illegible] on of [illegible] d im- [illegible] transmission line [illegible] pedance Z_r, is

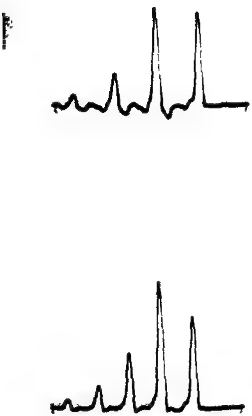


Fig 2. Typical transient phenomena in a transmission line. These are oscillographic recordings of voltage as a function of time at the sending end of a 300-meter transmission line with the receiving end open-circuited Time increases from right to left; the first (right-hand) pulse is delivered by a generator, equivalent to an open circuit, so that a new forward wave results from each reflected wave arriving at the sending end At the end of each 2-μsec interval, an echo arrives from receiving end. In the upper trace minor discontinuities in the line at intermediate points result in intermediate echos Intermediate discontinuities minimized in lower trace

$$Z_s = \frac{Z_r \cos \beta x + jZ_0 \sin \beta x}{\cos \beta x + j(Z_r / Z_0) \sin \beta x}$$

This equation describes the property of a length of line which transforms an impedance Z_r to a new impedance Z_s. In the simple cases, in which Z_r is a short circuit or open circuit, Z_s is a reactance Various lengths of line may be used to replace more conventional capacitors or inductors These properties are widely applied at high frequencies, where suitable values of βx require only physically short lengths of line.

Tapered transmission lines. Transmission lines with progressively increasing or decreasing spacing are used as impedance transformers at very high frequencies and as pulse transformers for pulses of millimicrosecond duration Although tapers designed to produce exponential-varying parameters, as in the exponential line, are most common, a number of other tapers are useful *See* MICROWAVE TRANSMISSION LINES. [F.M.W.]

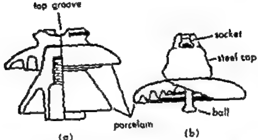


Fig. 6 Typical insulators used on transmission lines (a) Large two-piece pin-type insulator. (b) One unit of a suspension insulator (From H Pender and W. A Del Mar, *Electrical Engineers' Handbook*, Wiley, 1949)

bare, stranded, hard-drawn copper or aluminum. Steel reinforced aluminum cable (ACSR) has a steel core for providing the requisite tensile strength If the conductor diameter required for low corona loss is greater than that of an ordinary stranded conductor having the cross sectional area required for the desired resistance, special hollow conductors are used (Fig 7). Some very high voltage lines use multiple, or bundle, conductors, each consisting of several stranded conductors joined through metal spacers and hung from the same insulators

Splices in large conductors are usually made with metal sleeves squeezed over the butted ends of the conductor by hydraulic jacks

Sag and tension Conductors between adjacent supports hang in a curve called a catenary (Fig 8). For a given length of span, the greater the tension in the conductor, the smaller is the sag High mechanical tension is desirable to reduce sag and thus to permit use of longer spans or shorter towers while maintaining adequate ground clearance However, the tension must not exceed the tensile yield strength of the conductor under the worst condition, which occurs under a combination of low temperature (causing shortening) simultaneously with the thickest coating of ice on the conductor and the strongest wind

Vibration At times the wind causes the conductors to vibrate with low amplitude and audible frequency This vibration bends the conductor where it is clamped to the insulators and eventually may produce fatigue breakage

A device with duplex weights is used on some lines to reduce conductor vibration. One or more of these dampers are fastened to the conductors several feet from the insulator clamp (Fig 9).

Sleet An ice covered conductor acts as an air foil and is lifted by the wind so that "dancing" occurs, of such amplitude that one conductor may strike another, producing a short circuit. Conductors should be so located that contact will not be made. Formation of ice is prevented if the current heats the conductor sufficiently. Some power companies make a practice of periodically taking endangered lines out of service and sending high currents through them

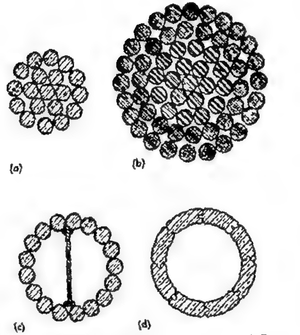


Fig 7. Typical transmission-line conductors (a) Copper conductor of 19 strands (b) Steel-reinforced aluminum cable (ACSR) with 19 steel and 42 aluminum strands (c) Hollow copper conductor with I-beam core. (d) Type HH hollow copper conductor.

Corona When the voltage gradient, or electric field strength, at the surface of the conductor exceeds the breakdown gradient of air, the air near the conductor surface becomes ionized. This condition, called corona, is evidenced by a visible glow at night and by a buzzing noise. *See* CORONA DISCHARGE.

Corona represents a loss of power and interference with radio reception, both of which increase

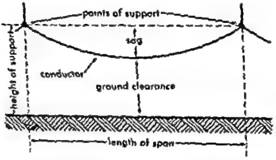


Fig 8 Catenary curve assumed by one span of a flexible conductor with supports at equal elevations

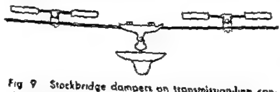


Fig 9 Stockbridge dampers on transmission-line conductor (From H Pender and W A Del Mar, *Electrical Engineers' Handbook*, Wiley, 1949)

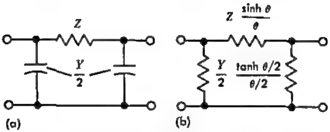


Fig. 4. Lumped-constant representations of a single-phase transmission line or of one phase of a three-phase line (a) Nominal π. (b) Equivalent π

along the line, is assumed to be lumped and divided into two equal parts, one at each end of the line. For a long line the theoretically exact equivalent π circuit should be used. Each of its branches is calculated by multiplying the corresponding branch of the nominal π by a correction factor given in Fig. 4*b*

By use of equivalent circuits, the steady-state electrical performance of a line can be calculated by the ordinary theory of ac circuits with lumped constants. These circuits can be combined with circuits representing series capacitors, transformers, and loads. If a complicated network is to be studied, it can be represented by a low-power model in a network analyzer wherein each line is represented by its π circuit

Aerial transmission lines. An aerial line consists of a set of conductors, usually bare, which is supported at a specified distance apart, and with a clearance or specified distance above ground surface. This method is known as open-wire construction. Cable installation for high-voltage aerial lines is relatively costly.

Routes. Lower-voltage transmission lines are usually built along highways, whereas higher-voltage lines are put on a special right of way, cleared of trees and brush Such routes are often chosen from results of aerial surveys.

Supporting structures. Lower-voltage aerial lines are usually supported by wooden poles and higher-voltage lines by wooden H frames or steel towers Rigid steel towers give the greatest strength and reliability. The higher the voltage, the greater must be the spacing between conductors and the clearance from conductor to ground. The farther apart the towers are placed, the greater is the sag of the conductors and the taller and stronger the towers must be. Fig. 5 shows some typical structures.

Insulators. Conductor supports, or insulators (Fig 6), are generally made of glazed porcelain. On lower-voltage lines, they are usually of the pin or post type On higher-voltage lines, they are of the suspension type, consisting of several units connected by swivel joints. The number of units per string depends on the desired impulse flashover voltage, but is not proportional to it, because the voltage does not divide equally between the several units.

Conductors These are wires suitable for carrying electric current. They are usually made of

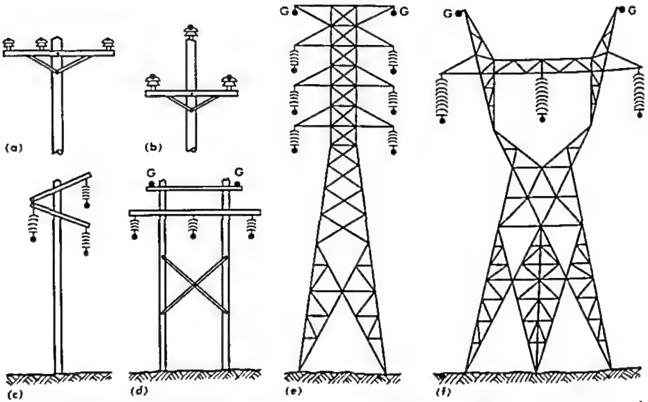


Fig. 5. Typical supporting structures used for electric-power transmission lines· (a-d) Wood poles; (e, f) Steel towers Structures in (a-d,f) are for single-circuit lines, (e) is for double-circuit lines In (a,b) pin-type insulators are used; in (c-f) suspension insulators are used Structures in (d-f) have ground wires (G) above the line conductors

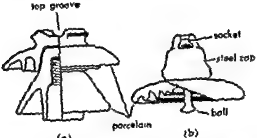


Fig. 6. Typical insulators used on transmission lines. (a) Large two-piece pin-type insulator. (b) One unit of a suspension insulator. (From H. Pender and W. A. Del Mar, Electrical Engineers' Handbook, Wiley, 1949)

bare, stranded, hard-drawn copper or aluminum. Steel-reinforced aluminum cable (ACSR) has a steel core for providing the requisite tensile strength. If the conductor diameter required for low corona loss is greater than that of an ordinary stranded conductor having the cross-sectional area required for the desired resistance, special hollow conductors are used (Fig. 7). Some very high voltage lines use multiple, or bundle, conductors, each consisting of several stranded conductors joined through metal spacers and hung from the same insulators.

Splices in large conductors are usually made with metal sleeves squeezed over the butted ends of the conductor by hydraulic jacks.

Sag and tension. Conductors between adjacent supports hang in a curve called a catenary (Fig. 8). For a given length of span, the greater the tension in the conductor, the smaller is the sag. High mechanical tension is desirable to reduce sag and thus to permit use of longer spans or shorter towers, while maintaining adequate ground clearance. However, the tension must not exceed the tensile yield strength of the conductor under the worst condition, which occurs under a combination of low temperature (causing shortening) simultaneously with the thickest coating of ice on the conductor and the strongest wind.

Vibration. At times the wind causes the conductors to vibrate with low amplitude and audible frequency. This vibration bends the conductor where it is clamped to the insulators and eventually may produce fatigue breakage.

A device with duplex weights is used on some lines to reduce conductor vibration. One or more of these dampers are fastened to the conductors several feet from the insulator clamp (Fig. 9).

Sleet. An ice-covered conductor acts as an air foil and is lifted by the wind, so that "dancing" occurs, of such amplitude that one conductor may strike another, producing a short circuit. Conductors should be so located that contact will not be made. Formation of ice is prevented if the current heats the conductor sufficiently. Some power companies make a practice of periodically taking endangered lines out of service and sending high currents through them.

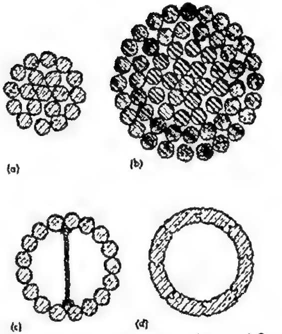


Fig. 7. Typical transmission-line conductors. (a) Copper conductor of 19 strands. (b) Steel-reinforced aluminum cable (ACSR) with 19 steel and 42 aluminum strands. (c) Hollow copper conductor with I-beam core. (d) Type HH hollow copper conductor.

Corona. When the voltage gradient, or electric field strength, at the surface of the conductor exceeds the breakdown gradient of air, the air near the conductor surface becomes ionized. This condition, called corona, is evidenced by a visible glow at night and by a buzzing noise. See CORONA DISCHARGE.

Corona represents a loss of power and interference with radio reception, both of which increase

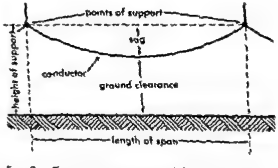


Fig. 8. Catenary curve assumed by one span of a flexible conductor with supports at equal elevations.

Fig. 9. Stockbridge dampers on transmission-line conductor. (From H. Pender and W. A. Del Mar, Electrical Engineers' Handbook, Wiley, 1949)

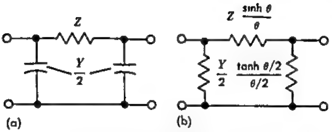


Fig. 4. Lumped-constant representations of a single-phase transmission line or of one phase of a three-phase line. *(a) Nominal π. (b) Equivalent π.*

along the line, is assumed to be lumped and divided into two equal parts, one at each end of the line. For a long line the theoretically exact equivalent π circuit should be used. Each of its branches is calculated by multiplying the corresponding branch of the nominal π by a correction factor given in Fig. 4*b*.

By use of equivalent circuits, the steady-state electrical performance of a line can be calculated by the ordinary theory of ac circuits with lumped constants. These circuits can be combined with circuits representing series capacitors, transformers, and loads. If a complicated network is to be studied, it can be represented by a low-power model in a network analyzer wherein each line is represented by its π circuit.

Aerial transmission lines. An aerial line consists of a set of conductors, usually bare, which is supported at a specified distance apart, and with a clearance or specified distance above ground surface. This method is known as open-wire construction. Cable installation for high-voltage aerial lines is relatively costly.

Routes. Lower-voltage transmission lines are usually built along highways, whereas higher-voltage lines are put on a special right of way, cleared of trees and brush. Such routes are often chosen from results of aerial surveys

Supporting structures. Lower-voltage aerial lines are usually supported by wooden poles and higher-voltage lines by wooden H frames or steel towers. Rigid steel towers give the greatest strength and reliability. The higher the voltage, the greater must be the spacing between conductors and the clearance from conductor to ground. The farther apart the towers are placed, the greater is the sag of the conductors and the taller and stronger the towers must be. Fig. 5 shows some typical structures

Insulators. Conductor supports, or insulators (Fig. 6), are generally made of glazed porcelain On lower-voltage lines, they are usually of the pin or post type. On higher-voltage lines, they are of the suspension type, consisting of several units connected by swivel joints. The number of units per string depends on the desired impulse flashover voltage, but is not proportional to it, because the voltage does not divide equally between the several units

Conductors These are wires suitable for carrying electric current. They are usually made of

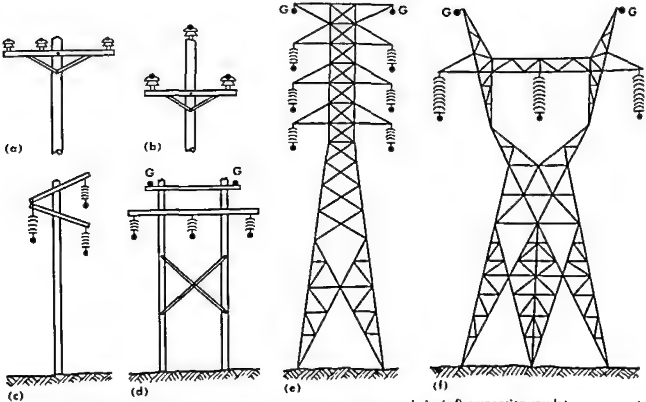


Fig. 5 Typical supporting structures used for electric-power transmission lines· (a-d) Wood poles; (e, f) Steel towers. Structures in (a-d,f) are for single-circuit lines; (e) is for double-circuit lines In (a,b) pin-type insulators are used; in (c-f) suspension insulators are used Structures in (d-f) have ground wires (G) above the line conductors

the order of 10^{10} cps, and the frequency of visible light is of the order of 5×10^{14} cps. The techniques of generating electromagnetic waves depend upon the frequency being used, as do the techniques of transmitting the energy to another location and of utilizing it once received.

The communication of information to a distant point is generally accomplished through the use of electromagnetic energy as a carrier. A familiar example is the telephone, in which sound waves in the range of frequencies from a few hundred to a few thousand cycles per second are converted into electromagnetic waves of the same frequencies, and these waves are guided by a pair of wires to their destination. Another familiar example is radio, in which the signals are caused to modify an identifiable characteristic, such as the amplitude or the frequency, of an electromagnetic carrier wave. The electromagnetic wave, thus modified, or modulated, is radiated from an antenna and can be received over a considerable region. See MODULATION; RADIO BROADCASTING; TELEPHONY.

Characteristics of electromagnetic waves. Figure 1 illustrates schematically some of the essential features of an electromagnetic wave. Shown in this figure are vectors which represent the electric field intensity E and the magnetic field intensity H at various points along a straight line taken in the direction of propagation of the wave. In this illustration the electric field is in a vertical plane and, at a given instant, varies ...

tion o

The n

mal to that of E and is, at each point, proportional in magnitude to E such that

$$\frac{E}{H} = \sqrt{\frac{\mu}{\epsilon}} \tag{1}$$

where H is the magnetic field intensity in amp/m, E is the electric field intensity in volts/m, ϵ is the permittivity, or absolute dielectric constant, of the medium, and μ is the absolute permeability of the medium. For a vacuum, $\epsilon = 8.854 \times 10^{-12}$ farad/m and $\mu = 4\pi \times 10^{-7}$ henry/m; therefore the ratio E/H is approximately 377 ohms. This ratio is termed the wave impedance of the medium. See MAXWELL'S EQUATIONS; WAVE EQUATION.

The E ...

sug

of

de

thr

... the axis and turned from E toward H, the longitudinal direction of travel of the screw is the direction of propagation of the energy.

The velocity of travel of the wave is

$$v = 1/\sqrt{\mu\epsilon} \tag{2}$$

In a vacuum, this is approximately 3×10^{8} meters/sec. The velocity in air is only slightly smaller.

The wavelength is the distance between two successive similar points on the wave, measured along the direction of propagation. The wavelength is denoted by λ in Fig. 1.

Fig. 1. Representation of an electromagnetic wave at a particular instant of time.

As the wave travels past a stationary point, the values of E and H at the point vary sinusoidally with time. The time required for one cycle of this variation is termed the period, T seconds. The number of cycles per second is the frequency f; and f equals $1/T$. In one cycle, the wave, traveling at the velocity v, moves one wavelength along the axis of propagation. Therefore,

$$\lambda = v/f \tag{3}$$

Assuming a velocity of 3×10^{8} m/sec, an electromagnetic wave having a frequency of 60 cps has a wavelength of 5×10^{6} meters (m), or approximately 3100 miles (mi). At 3 megacycles per second (Mc) λ is 100 m, and at 3000 Mc, λ is 10 centimeters (cm). Visible light has a frequency of the order of 5×10^{14} cps and a wavelength of approximately 6×10^{-5} cm.

The density of energy of an electric field is $\epsilon E^2/2$ joules/m³, and that of a magnetic field is $\mu H^2/2$ joules/m³. With the aid of Eq. (1),

$$\mu H^2/2 = \mu(\sqrt{\epsilon/\mu}\,E)^2/2 = \epsilon E^2/2$$

Therefore, in an electromagnetic wave, the electric and magnetic fields carry equal energies. The total energy density at any point is equal to ϵE^2 joules/m³. Since this is transported with a velocity equal to $1/\sqrt{\mu\epsilon}$, the rate of flow of energy per square meter normal to the direction of propagation, is $\epsilon E^2 v$ or $E^2\sqrt{\epsilon/\mu}$ watts/m². In radio broadcasting, a field strength of 50 mv/m is considered to be strong. An electromagnetic wave with this intensity has an average energy density of only 2.2×10^{-14} joule/m³, and the average rate of energy flow per square meter is 6.6×10^{-6} watt/m².

Radiation from an antenna. Figure 2 illustrates the configuration of the electric and magnetic fields about a short vertical antenna in which flows a sinusoidal current of the form $i = I_{max} \sin 2\pi ft$ amp. The picture applies both to an antenna in free space ...

upper

at the

In the

to a fi

The fi... have symmetry about the axis through the antenna. For pictorial simplicity only selected portions of the fields are shown in Fig. 2. The magnetic field is circular about the antenna, is perpendicular at every point to the direction of the electric field, and is proportional in intensity to the

rapidly with voltage Transmission lines are normally operated at a voltage below that at which corona becomes appreciable. The larger the conductor diameter, the higher the operating voltage may be. Large conductors (1–2 in.) are required for high voltages, such as 220 and 330 kv.

Inductive coordination. If a telephone line runs near and parallel to a power line for some distance, the high currents and voltages in the power line may induce currents and voltages in the telephone line These signals may be comparable in strength to the telephone signals and thus produce objectionable noise in telephone receivers. The coupling between the two lines can be reduced by transposition of either, or preferably both. lines (see Fig. 10) and by shielding, such as that provided by ground wires and grounded cable sheaths. Special towers are required at transposition points of the power line. *See* COMMUNICATIONS SYSTEMS PROTECTION.

Inspection and fault location. Transmission lines should have both periodic general inspections and special immediate inspections of points where short circuits have occurred, to detect damage, such as broken insulators, which might impair the reliability of the line Faults can be located approximately by electrical measurements made from the ends of the line and then exactly by visual patrol of the vicinity. *See* CIRCUIT TESTING, ELECTRICAL

Lightning protection Lightning is the most detrimental factor affecting the reliability of electric power service, but its damaging effects have been greatly reduced by proper design (*see* LIGHTNING AND SURGE PROTECTION) Lightning striking a transmission line *momentarily impresses a very* high voltage on the line, causing spark-over to ground, usually at an insulator Power current then follows the spark path, producing an arc, which constitutes a short circuit and which can be extinguished only by disconnecting the faulted line from the rest of the power network Lines built where severe thunderstorms are prevalent are equipped with overhead ground wires (Fig. 5*d, e, f*) for intercepting the lightning stroke and leading it to ground at the nearest tower.

Underground and submarine lines. Insulated cables are used in congested areas where the cost of right of way for overhead lines would be ex*cessive, in city streets where overhead lines would* be too unsightly or hazardous, in and around power stations, and for crossing wide bodies of water. Cables rated up to 69 kv are common, with occasional ratings up to 230 kv. Both single-conductor and three-conductor cables are used.

High-voltage cables are constructed with stranded annealed copper conductors, usually insulated by wrappings of many layers of thin paper tape impregnated with viscous mineral oil. One or more conductors thus insulated are enclosed in an extruded lead sheath for excluding moisture. Some high-voltage cables employ fluid oil or inert gas under pressure to prevent the formation of voids in the solid insulation

Although some cables are buried directly in the earth, most of them are placed in buried banks of fiber or tile ducts running between manholes where the cable splices are made. Thus the cables are protected, and it is possible to replace a damaged section of cable without excavation.

The power that a cable can carry is limited by its temperature rise. Cables have lower inductance and higher capacitance and require more capacitive charging current than aerial lines of the same length and voltage. [F W.K.]

Bibliography: Electrical Transmission and Distribution Reference Book, 4th ed., 1950; R. W. P. King, *Transmission-Line Theory*, 1955; A E Knowlton (ed), *Standard Handbook for Electrical Engineers*, 9th ed., 1957; H Pender and W A. Del Mar (eds), *Electrical Engineers' Handbook*, 4th ed., 1949; E. R Schatz and E M Williams, Pulse transients in exponential transmission lines, *Proc IRE*, vol 38, pt. 2, 1950; E M Williams and J B Woodford, Jr., *Transmission Circuits*, 1957

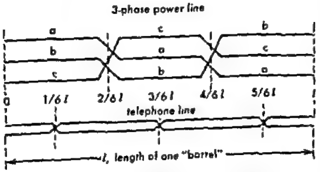


Fig. 10 Transposition scheme of three-phase power line and one telephone circuit.

Transmission theory and methods

The transmission of electrical energy by wires, the broadcasting of radio signals, and the phenomenon of visible light are all examples of the propagation of electromagnetic energy Electromagnetic energy travels in the form of a wave Its speed of travel in a vacuum is approximately 3×10^8 meters per second (186,000 miles per second) Its speed is slightly less in air and is considerably less in liquid and solid insulators, such as oil or polystyrene. An electromagnetic wave does not penetrate far into an electrical conductor, such as sea water or a metal, and a wave incident on such a surface is largely reflected

Electromagnetic waves originate from accelerated electric charges. A radio wave originates from the accelerated electrons in the sending antenna, and a light wave originates when electrons fall from one energy level to another in an atom The waves emitted from a source are generally oscillatory in character and are described in terms of their frequency of oscillation. Local telephone lines (not using carrier systems) carry electromagnetic waves with frequencies of about 300–4000 cycles per second (cps), broadcast radio uses frequencies of the order of 10^6 cps, radar uses frequencies of

tenna, but is oversimplified insofar as communication to and from positions on or near the Earth is concerned. The ground is a reasonably good, but not perfect, conductor; hence, the actual propagation over the surface of the Earth will show a more rapid decrease of field strength than that indicated by the factor of $1/r$ in Eq. (4). Irregularities and obstructions may interfere. For long-range transmission, the spherical shape of the Earth is important. Inhomogeneities in the atmosphere refract the wave somewhat. For long range transmission, the ionized region high in the atmosphere known as the Kennelly-Heaviside layer, or ionosphere, can act as a reflector. The electric field of the wave produces oscillation of the charged particles of the region and this causes the refractive index of the layer to be smaller than that of the atmosphere below. The result is that, if the angle of incidence is not too near the normal and if the frequency of the wave is not too high, the wave may be refracted back toward the Earth. Successive reflections between ionosphere and Earth can provide communication for long distances around the periphery of the Earth. *See* RADIO-WAVE PROPAGATION

Hollow wave guides. When an electromagnetic wave is introduced into the interior of a hollow metallic pipe of suitably large cross-sectional dimensions, the energy is guided along the interior of the pipe with comparatively little loss. The most common cross-sectional shapes are the rectangle and the circle. The cross-sectional dimensions of the tube must be greater than ...

the wavel

gate in th

are commonly used only at wavelengths of 10 cm or less (frequencies of 3000 Mc or higher).

A single wave of the type in Fig. 1 cannot propagate longitudinally in a tubular conductor since, at some portions of the inner surface of the conducting tube, the E vector of the wave necessarily has a component tangential to the surface. This is impossible because an electric field cannot be established along a good conductor such as the wall of the tube. An electromagnetic wave can propagate along the interior of the tube only by reflecting back and forth between the walls of the tube. This reflection is a comparatively simple one between the plane surfaces of a rectangular tube but is a complex reflection in tubes of other cross-sectional shapes.

A dielectric rod can be used in a similar fashion as a wave guide. Such a rod, if of sufficient cross-sectional dimensions, can contain the electromagnetic wave by the phenomenon of total reflection at the surface.

A hollow metallic wave guide of rectangular cross section is shown in Fig. 3a. The simplest mode of propagation is indicated in Fig. 3b. The entire space is filled with a plane electromagnetic wave which moves obliquely to the ...

... normal to the direction of propagation is a plane of equal phase (thus the name plane wave), and one such plane is indicated in the illustration by a dashed line. The wave strikes the wall at an angle θ from the normal and is reflected at an equal angle. As the wave is reflected, the direction of its E vector is reversed so as to make the tangential component of the electric field equal to zero at the conducting wall. The wave incident on the left wall thus is reflected to the right, where it is again reflected and moves to the left. By successive reflections the energy propagates longitudinally along the interior of the guide. As the wave incident upon the wall reflects and reverses the direction of its E vector, electric currents are caused to flow in the conducting wall. Since the wall is not a perfect conductor, some of the energy of the wave is transformed into heat. Consequently the amplitude of the wave diminishes exponentially as it passes down the guide; this phenomenon is termed attenuation. For an electromagnetic wave with a frequency of 3000 Mc (wavelength of 10 cm) propagating down the interior of a rectangular copper wave guide with cross-sectional dimensions of 1.5 in. by 3 in., half the power is lost in a length of approximately 500 ft. Hollow wave guides are used chiefly for short-distance transmission, as from a transmitter to an antenna.

The requirements on the reflection of the wave, as outlined above, restrict the wavelength that can be propagated in a hollow guide. Consider the ray ABC in Fig. 3. The wave propagates from A to B, where it is reflected with reversal of the E vector; thereupon it propagates from B to C, where it is again reflected with another reversal of the E vector. But AC is a line of equal phase, and so the wave emerging from C must have the same phase as that at A. Thus the distance ABC must be an

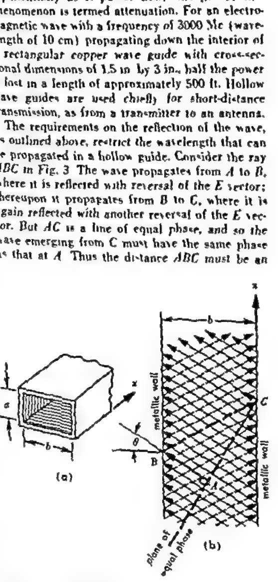


Fig. 3. A hollow metallic wave guide of rectangular cross section. (a) The guide. (b) Paths of electromagnetic energy in the simplest mode of propagation.

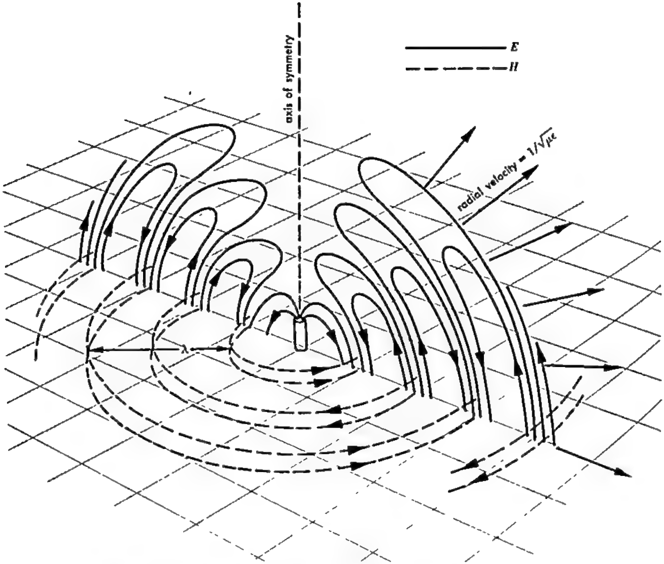


Fig 2. Configuration of electric and magnetic fields about an antenna

magnitude of the electric field, as indicated by Eq (1). All parts of the wave travel radially outward from the antenna with the velocity given by Eq (2); the wave is described as spherical, with the antenna at the center. The wavelength of the radiation is given by Eq. (3).

If a short antenna projecting above a highly conducting plane surface carries a current of $i = I_{max} \sin 2\pi ft$ that is uniform throughout the length of the antenna, the intensity of the electric field in the radiated wave is

$$E_{max} = \sqrt{\frac{\mu}{\epsilon}} \frac{I_{max}}{r} \frac{l}{\lambda} \cos\theta \quad \text{volts/m} \qquad (4)$$

where l is the length of the antenna, r is the radial distance from the antenna, θ is the angle measured from the horizontal The radiated field intensity is zero directly above the antenna, is greatest along the conducting plane where θ is 0, and varies inversely with the distance from the antenna.

If the rate of flow of energy per unit area $E^2\sqrt{\epsilon/\mu}$ is integrated over an imaginary spherical surface about the antenna, the following result is obtained for the average power radiated

$$P_{av} = \frac{2\pi}{3} \sqrt{\frac{\mu}{\epsilon}} I^2_{max} \left(\frac{l}{\lambda}\right)^2 \quad \text{watts} \qquad (5)$$

The factor $(l/\lambda)^2$ is of particular importance, for it indicates that a longer antenna is required at the longer wavelengths (lower frequencies). The radiation of appreciable energy at a very low frequency requires an impracticably long antenna

The foregoing relations assume a uniform current throughout the length of the antenna, an approximation to this can be achieved in practice by connecting a long horizontal conductor to the top of the antenna Where some such construction is not utilized, the current will not be uniform in the antenna and will, in fact, be zero at the tip For further discussion, see ANTENNA (AERIAL)

It is possible to construct directive antennas which, by the addition of more antenna elements and reflectors, direct most of the radiation in a particular direction

Propagation over the Earth. The foregoing discussion shows some of the important features of the radiation of electromagnetic energy from an an-

in a distance of perhaps 60 miles. The losses increase with frequency, and for a typical air-insulated coaxial line operating at 5 Mc, half the energy is lost in a distance of less than a mile. At a frequency of 3000 Mc ($\lambda = 10$ cm), typical distances in which half the energy is lost are, for air-insulated coaxial cable, 80 ft; for flexible coaxial cable insulated with polyethylene, 30 ft.

Noise. In a transmission line intended for the transmission of large amounts of power, such as the cross-country lines joining electrical generating stations to centers of population, the loss of an appreciable proportion of the power en route is a serious matter. In a communication system, however, the average rate of flow of energy is rather small, and the intrinsic value of the energy itself is not of prime importance. The important characteristic of such a system is the accurate transmission of information, and the limiting factor is noise. Noise is always present in a transmission channel. Two common causes are thermal agitation and nearby electrical discharges. In a transmission system conveying information by an electromagnetic wave, the loss of energy in transmission becomes a serious matter if the wave is attenuated to the point where it is not large enough to override the noise. Amplifiers must be inserted in the transmission system at sufficiently close intervals so that the signal never falls into the noise level, from which it could not be recovered and interpreted accurately.

Circuit analysis of transmission lines. Because the conductors of a transmission line are almost always spaced much closer together than a quarter wavelength of the electromagnetic energy that they are guiding, it is possible to analyze their performance quantitatively by circuit theory. It is then possible to deal with the voltages between the conductors and the currents flowing along the conductors, instead of with the electric and magnetic fields that exist in the insulating medium. See TRANSMISSION LINES

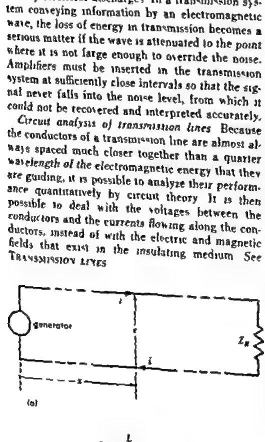


Fig. 5. Representation of a transmission line. (a) Circuit diagram. (b) Visualization of L and C.

The wave-guiding properties of the transmission line can be examined most conveniently if losses of energy are ignored. If L is defined as the inductance of the pair of conductors per unit length and C the capacitance between the conductors per unit length, field theory shows that L is μF_g and C is ϵ/F_g, where F_g is a geometrical factor that depends on the cross sectional configuration of the conductors. For a coaxial line (Fig. 4a), F_g is $(1/2\pi) \ln (b/a)$. For a two-wire line (Fig. 4b), F_g is approximately $(1/\pi) \log_e (D/a)$.

In the circuit analysis, the line can be visualized as composed of a cascaded set of sections, each of short length Δx, as shown in Fig. 5b. The partial differential equations which describe the voltage e and the current i are

$$\frac{\partial e}{\partial x} = -L\frac{\partial i}{\partial t} \qquad \frac{\partial i}{\partial x} = -C\frac{\partial e}{\partial t} \tag{6}$$

The solution of these equations is

$$e = f_1(x - t/\sqrt{LC}) + f_2(x + t/\sqrt{LC})$$
$$i = \frac{1}{\sqrt{L/C}}[f_1(x - t/\sqrt{LC}) - f_2(x + t/\sqrt{LC})] \tag{7}$$

where f_1 and f_2 are any finite, single-valued functions of the arguments $x - t/\sqrt{LC}$ and $x + t/\sqrt{LC}$ respectively. These are interpreted physically as traveling waves, the first traveling in the positive x direction with the speed $1/\sqrt{LC}$ and the second traveling in the negative x direction at the same speed. Substitution of the values for L and C for any configuration of conductors yields the velocity $1/\sqrt{LC}$, which equals $1/\sqrt{\mu\epsilon}$.

The quantity $\sqrt{L/C}$ has the dimensions of ohms, and is termed the characteristic impedance Z_0 of the line.

$$Z_0 = \sqrt{L/C} = \sqrt{\mu/\epsilon}\, F_g$$

Z_0 is a real quantity (a resistance) and is equal to the wave impedance of the insulating medium [Eq. (1)] multiplied by the geometrical factor characteristic of the particular configuration of conductors. For the traveling waves of voltage and current of Eq. (7), the ratio of voltage to current of the forward-traveling wave is Z_0; that of the backward-traveling wave is $-Z_0$.

In Fig. 5a, a source of electrical energy is connected at one end of a transmission line and an electrical load is connected to the other. Electromagnetic energy is propagated from the sending end to the receiving end; a portion of the energy is reflected back toward the sending end if the load impedance Z_R is different from the characteristic impedance Z_0 of the line. If Z_R equals Z_0, there is no reflection of energy at the load, and, in Eq. (7),

integral multiple of a wavelength, or $n\lambda$, where n is a positive integer. The distance ABC is $2b \cos \theta$ where b is the breadth of the guide; hence, $n\lambda = 2b \cos \theta$. The condition for propagation down the axis of the guide is that $\theta > 0$; hence $\cos \theta < 1$, and the restriction on wavelength is $\lambda < 2b/n$. The greatest ratio of wavelength to breadth of guide is obtained when $n = 1$, whence $\lambda/b < 2$. Thus, the breadth of the guide must be *somewhat greater than* $\lambda/2$

In the simple mode of propagation described above, the fields are independent of distance in the direction of the dimension a, and this dimension has no influence on the propagation. The net electric vector caused by the sum of the two waves is everywhere transverse to the longitudinal axis of the guide, and so the mode is described as transverse electric (TE).

If the wavelength of the radiation is small enough in comparison with the cross-sectional dimensions of the guide, more complex modes of propagation are possible, in which the wave reflects obliquely against a side wall, proceeds to the top of the guide, and reflects from there to the other side wall, then to the bottom wall, and so on. With this type of reflection it is possible to have both transverse-electric and transverse-magnetic modes. In the latter the net H vector is everywhere transverse to the axis of the guide.

When the dimensions of the guide are such that complex modes are possible, so also are the simple ones The transmission of energy by a combination of modes introduces complications in abstracting the energy from the guide at the receiving end. Propagation in only the simplest mode is insured by selecting the dimension b to be greater than $\lambda/2$ but not as large as λ, and by restricting the dimension a so as to render complex modes impossible.

Wave guides of circular cross section are sometimes used. Analysis of these shows that the first TE mode is propagated if the diameter of the guide is greater than 0.586λ, and that the first TM mode is propagated if the radius is greater than 0.766λ. *See* WAVE GUIDE.

Two-conductor transmission lines. Electromagnetic energy can be propagated in a simple mode along two parallel conductors Such a wave-guiding system is commonly termed a transmission line. Two common forms are shown in Fig. 4 If the spacing between conductors is a small fraction of the wavelength of the transmitted energy, only one mode of propagation is possible This corresponds to the wave of Fig. 1, with the direction of propagation taken longitudinally along the line. The E and H vectors are in the plane of the cross section, and the mode is termed transverse electromagnetic (TEM). The E vector must be at right angles to a highly conducting surface, and the oscillating H vector must be parallel with such a surface. With two separated conductors, there is *for each geometrical arrangement of conductors one and only one cross-sectional field configuration* which will satisfy the boundary conditions at the metal surface. The field configurations for coaxial and two-wire lines are shown in Fig 4. At each point the ratio of E to H is as given by Eq. (1), and the velocity of propagation of the wave is as given by Eq (2). Half of the propagated energy is contained in the electric field and half in the magnetic field. This mode of propagation is in contrast with the more complex modes required in a hollow metal pipe, where the conditions required at the boundaries can be satisfied only by means of reflections at the metal walls. As a result, the two-conductor transmission line does not have the upper limit on wavelength that was imposed on the hollow wave guide by the requirement of reflections; in fact, the two-conductor line operates completely normally at zero frequency (direct current).

At wavelengths that are small enough to be comparable with the cross-sectional dimensions of the line, *more complex modes, involving reflections* from the surfaces of the conductors, become possible High-frequency energy can thus be propagated in several modes simultaneously. In a coaxial cable a rough criterion for the elimination of higher modes is that the wavelength should be greater than the average of the circumferences of the inner and outer conductors.

*As the wave propagates along the line, it is ac*companied by currents which flow longitudinally in the conductors These currents can be regarded as satisfying the boundary condition for the tangential H field at the surface of the conductor. The conductors have a finite conductivity, and so these currents cause a transformation of electrical energy into heat The energy lost comes from the *stored energy of the wave, and so the wave, as it* progresses, diminishes in amplitude The conductors are necessarily supported by insulators which are imperfect and cause additional attenuation of the wave In a typical open-wire telephone line operating at voice frequencies, half the energy is lost

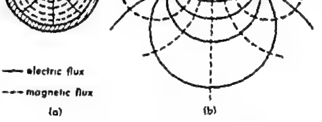


Fig. 4 Cross sections of common two-conductor transmission lines. (a) Coaxial cable. (b) Two-wire line.

nels This can be done, for example, by a switch with n contacts, each leading to an output channel, and a wiper which rotates synchronously with the one at the sending end. Once each frame the portion of the composite signal representing the sample of signal 1 is connected to channel 1 of the output, slightly later the portion of the composite signal representing the signal of channel 2 is connected to channel 2 of the output, and so on through the frame, whereupon the process repeats. The switch at the receiving end can be kept in proper synchronism by marking the beginning of each frame with a distinctive pulse, either larger than an ordinary signal or longer in duration. These pulses, by the use of appropriate circuitry, serve to keep the switching operation at the output in synchronism with that at the input.

Multiplexing by frequency division or time division, or by a combination of the two, is used in the transmission of data such as temperature as sensed by thermocouples, deflections as sensed by strain gages velocity and acceleration as sensed by appropriate transducers, and so on This process is termed telemetering *See* TELEMETERING.

The transmission of television pictures is accomplished by a time-division scheme in which the relative brightnesses of various portions of an image are sampled in time sequence at the sending end and are registered in the same sequence on a screen at the receiving end. The sampling at the sending end is done by scanning the image along a horizontal line, then moving vertically to a new line slightly displaced from the first and scanning again and so on until the whole image has been scanned A short synchronizing pulse is transmitted to mark the beginning of each line, and a long synchronizing pulse is used to mark the beginning of each frame *See* TELEVISION [W.C.JO.]

Bibliography. W. C. Johnson *Transmission Lines and Networks*, 1950, E C Jordan, *Electromagnetic Waves and Radiating Systems*, 1950; I. I. Karakash, *Transmission Lines and Filter Networks*, 1950, J D Kraus, *Electromagnetics*, 1953 S Ramo and J R. Whinnery *Fields* ... *Circuit Analysis of Transmission Lines*, 1958

Transmutation

The change of one element into another, either naturally, in radioactive elements or artificially, by bombarding an element with electrons, deuterons, or α-particles in particle accelerators, or with neutrons in atomic piles

Natural transmutation was first explained by Marie Curie about 1900 as the result of the decay of radioactive elements into others of lower atomic weight Ernest Rutherford produced the first artificial transmutation (nitrogen into oxygen and hydrogen) in 1919 Artificial transmutation is the method of origin of the heavier, artificial transuranium elements, and also of hundreds of radioactive isotopes of most of the 96 more common elements in the periodic table. Practically all of these ...

[C.CH.]

Transonic flight

In aerodynamics, flight of a vehicle at speeds near the speed of sound When the velocity of an airplane approaches the speed of ...

speeds The drag increases greatly, the lift at a given attitude decreases, the moments acting on the airplane change abruptly, and the vehicle may shake or buffet. Such phenomena usually persist to flight velocities somewhat above the speed of sound These flight characteristics, as well as the speeds at which they occur, are usually referred to as transonic The extent of the speed range of these changes depends on the form of the airplane; for configurations designed for subsonic flight they may occur at velocities of 70–110% of the speed of sound (Mach numbers of 0.7–1.1); for airplanes intended for transonic or supersonic flight they may be present only at Mach numbers of 0.95–1.05. *See* MACH NUMBER

Shock waves. The transonic flight characteristics result from the development of shock waves about the airplane (*see* SHOCK WAVES). Because of the accelerations of airflow over the various surfaces, the local velocities become supersonic while the airplane itself is still subsonic. (The flight speed at which such local supersonic flows first occur is called the critical speed.) Shock waves are associated with deceleration of these local supersonic flows to subsonic flight velocities. The usual form of these waves at speeds just above the critical values is illustrated by Fig 1, in which the change of pressure in the flow about an airfoil section is made visible (*see* SCHLIEREN PHOTOGRAPHY). The shock wave is the nearly vertical line near the midchord of the section. Such shock waves in the position shown cause abrupt streamwise increases of pressure on the airplane surfaces. These gradients may cause a reversal and separation of the flow in the boundary layer on the wing surface in roughly the same manner as do similar pressure changes at lower subcritical speeds (*see* BOUNDARY-LAYER FLOW) When the wing carries lift, the shock-induced separation is particularly strong on the upper surfaces

As ...
the fl...
of dra...
mome... separated flow results in an irregular change of the aerodynamic forces acting on the airplane with resultant buffeting and shaking As the Mach number is increased, the shock waves move aft ...

the function f_2, representing leftward-traveling energy, is absent. This is the condition desired when the purpose of the line is to deliver energy from the source to the load. The sending-end impedance of the line is then equal to Z_0.

In addition to impedance matching to reduce reflections (echos) along a transmission line, it is also necessary to minimize signal distortion, which consists of amplitude and phase (delay) distortion If the line attenuation is frequency-dependent, then a signal consisting of a group of different-frequency components will undergo amplitude distortion due to the unequal attenuation of each component of the signal Similarly, if the velocity of propagation along the line is frequency-dependent, then a delay in phase of each component will result in associated phase distortion of the signal.

Signal distortion is minimized by the use of line loading, which is the addition of series impedances along the line and which is used to adjust the line parameters to obtain the so-called distortionless condition. Under distortionless operation the attenuation and velocity of propagation are independent of frequency. For a discussion of the distortionless line *see* TRANSMISSION LINES Instead of loading a line, one may employ equalizing circuits to compensate for the phase distortion along the line.

Short sections of transmission line are sometimes used to provide low-loss reactive impedances and resonant circuits at high frequencies. This is done by open-circuiting or short-circuiting the receiving end of the line to provide complete reflection of the incident energy. A short-circuited low-loss line provides a sending-end impedance of

$$Z_s = jZ_0 \tan (2\pi fl/v) \quad \text{ohms} \qquad (8)$$

When l equals $v/4f$, the line is a quarter wavelength long, the argument $(2\pi fl/v)$ of the tangent function in Eq. (8) is $\pi/2$, and Z_s approaches an infinite value. In actual practice, losses keep Z_s to a finite value However, at high frequencies the quarter wavelength is short and the losses are small, and such a short-circuited quarter-wave section can be used successfully as a low-loss insulator. Such a section is a resonant one and can be used as a substitute for a parallel-resonant LC circuit, for example, as a tank circuit for a high-frequency oscillator. At low frequencies the required quarter wavelength is so large that losses impair the performance; also the length becomes inconveniently great. At a frequency lower than $v/4l$, the sending-end impedance of the short-circuited line is inductive, and at frequencies between $v/4l$ and $v/2l$ the impedance is capacitive. This provides the possibility of using sections of short-circuited line as reactive elements in circuits.

Simultaneous transmission of messages. One channel of propagation of electromagnetic energy can be used for the simultaneous transmission of a number of messages by either frequency division or time division, or by a combination of the two This process is sometimes termed multiplexing.

Information is impressed on a sinusoidal carrier by a process termed modulation. In amplitude modulation, the amplitude of the carrier is varied in amplitude in accordance with the signal This gives rise to additional sinusoidal waves, termed sidebands, which accompany the carrier but which are of somewhat different frequency from the carrier. Sidebands are also present in phase and frequency modulation. Thus, the carrying of information requires a set of waves which occupy a finite band of frequencies. In principle, the more rapid the transmission of information, the greater the band width required. *See* INFORMATION THEORY.

Frequency-division multiplexing. In the broadcasting of information by radiation of a carrier and sidebands from an antenna, the frequency-division scheme is used Various transmitters radiate electromagnetic energy at different frequencies, and each requires a finite band of frequencies to convey information. This limits the number of neighboring transmitters that can broadcast in any given range of frequencies. A receiver selects transmitted messages by a band-pass filtering action, the purpose of which is to reject the carrier and sideband waves from all transmitters except the one to which the filter is tuned

Frequency division is also applied to the multiplexing of channels of communication which employ guided waves. It is, for example, the principle employed in carrier telephony, where a set of conductors guides carrier waves and their sidebands in the TEM mode, and, at the receiving end, band-pass filters separate out the various channels of communication.

Time-division multiplexing The time-division method of multiplexing employs the principle of sampling (*see* PULSE MODULATION) If a signal contains frequencies up to a certain frequency f_1, as for example speech, which contains essential frequencies up to perhaps 5000 cps, a set of samples of this signal taken $2f_1$ times per second will adequately represent the signal If there are n signals to be sent over the same channel of communication, the first signal is sampled briefly, then the second, and so on to the nth, then the sampling is repeated This can be done, for example, by a switch with n contacts and a rotating wiper. Each set of n samples is termed a frame. Within each frame the n samples of the n signals occur in a definite time sequence, and the frames follow one after another. The resulting set of samples comprise a new signal which contains all the pertinent information of the n original signals

The composite signal can be sent directly over a transmission facility such as a transmission line, or it can be used to modulate a sinusoidal wave which will act as a carrier The resulting carrier and its sidebands can be broadcast with an antenna or guided to its destination by an appropriate wave-guiding system. At the receiving end the wave is demodulated so as to recover the composite signal The n samples contained within each frame are then separated and directed into n output chan-

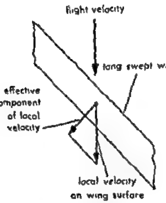


Fig 3 Effect of wing sweepback

used in practice, experience has indicated that excessive sweep leads to a number of aerodynamic problems. Most significantly, a highly swept wing may have an abrupt nose up moment at the higher lifts. This phenomenon, known as pitch-up, may result in excessive aerodynamic loads or stall. Pitch-up results from an initial separation of the boundary layer on the outboard part of the swept wing with an associated loss of lift for this region. Because this portion of a swept wing is aft of the center of gravity, the loss of lift causes a nose-up moment. Increasing the wing sweep also reduces the lift available for take off and landing.

Because of these limitations, most transport-type airplanes designed to fly at a speed close to the speed of sound incorporate only a moderate amount of wing sweep. Usually the obliqueness of the midchord element for such airplanes is about 30°. Such a sweep provides a delay of the onset of the adverse transonic characteristics of roughly 0.08 Mach number. Transonic and supersonic military airplanes may also incorporate as much as 45° of midchord sweep, which results in a delay of the adverse effects of approximately 0.15 Mach number.

Reduced thickness ratio. Substantial improvements in the adverse transonic characteristics are provided by reducing the thickness-to-chord ratios for the wing and tail surfaces. Such changes reduce the acceleration of the flow over the surfaces with a resulting delay in the onset of local supersonic flows and the associated shock wave. Also, the severity of the adverse longitudinal pressure gradients on the wing surface is lessened so that boundary layer separation is reduced. However, reductions of thickness lead to considerable increases in the weight of a structurally sound wing. Thus the wing thickness used must be a compromise between the aerodynamic and structural factors. Most high speed transport wings have mean thickness-to-chord ratios of about 10%, whereas military airplanes may have thickness ratios as low as 3%.

Reduced aspect ratio. Reductions in the aspect ratio provide delays and reductions of the transonic changes similar to those provided by reductions in thickness ratio, although the magnitude of the effect is usually considerably less. More important, lower aspect ratios result in improvements of the wing structural characteristics, which allow the use of smaller thickness-to-chord ratios. However, the use of reduced aspect ratios increases the subcritical drag due to lift. Most high-speed transport wings have aspect ratios of about 7; transonic and supersonic military airplanes may have aspect ratios as low as 2.

Added bodies. The adverse transonic characteristics may also be improved by adding streamlined bodies to the aft portion of the upper surface of the wing. Such changes provide reductions of the accelerated flows and adverse pressure gradients similar to those provided by reducing the thickness ratio.

Area rule. Because of the pronounced interaction of the shock waves of various airplane components near the speed of sound, the drag increase for the airplane associated with these waves is most effectively defined and improved by considering the flow about the configuration as a whole. For a nonlifting condition, the forms of the shock waves and consequently the drag are primarily a function of the longitudinal development of cross-sectional area, in section normal to the airstream, for the complete airplane. This relationship, called the area rule, is illustrated by considering that the transonic drag increases for the configurations shown in Fig. 4. The various normal cross-sectional areas for the body of revolution, such as at BB, are the same as those for the wing fuselage combination at the corresponding longitudinal station, such as at AA. The shock wave and the resulting drag near the speed of sound are approximately the same for the two configurations.

On the basis of the area rule, the transonic drag increment is reduced by shaping and arranging the airplane components so that area development for the airplane more nearly approaches the shape with the lowe...
magnitude ...
is greatly ...
and reduc...

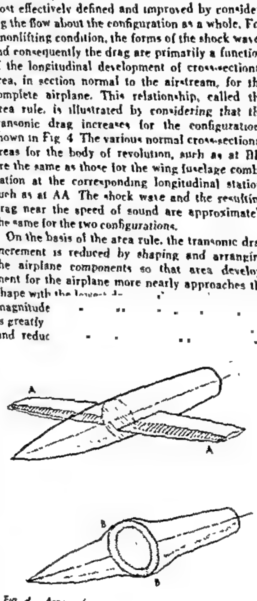


Fig 4 Area rule comparison

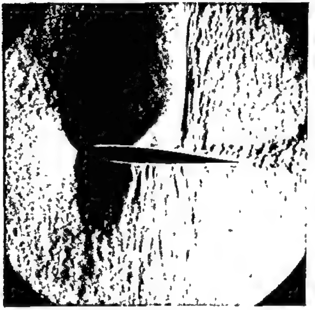


Fig. 1. Schlieren photograph of flow about airfoil section at low transonic speeds

surfaces. With the shocks in these positions, the associated pressure gradients have relatively little effect on the boundary layer, and the shock-induced separation is greatly reduced

Effects on flight characteristics. When the speed is increased to the higher transonic range, at and just above the speed of sound, the energy losses in the shock waves about an airplane may become large. As a result, the drag may increase to many times the subsonic value. At these speeds shock waves, in addition to those present near the aft parts of the surfaces, form ahead of the components. The various waves extend outward, interact, and merge to form two shock waves at a distance from the airplane (Fig 2). These two waves are relatively strong and extensive; they may extend for miles from the airplane. *See* SONIC BOOM.

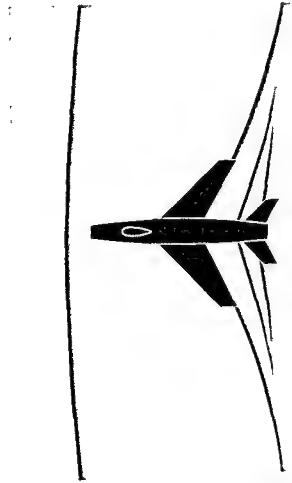


Fig 2 Shock waves about airplane near the speed of sound.

As the speed is increased through the transonic range, the changes of the distribution of load on the wing, resulting first from boundary-layer separation and then from rearward movement of the shock wave, cause a marked rearward shift of the center of lift. This shift causes a nose-downward moment on the airplane that must be corrected by an increase of the negative lift on the usual tail to maintain trim. Also, the effectiveness of the usual flap-type elevator and aileron control surfaces used on subsonic airplanes decreases greatly at transonic speeds At subcritical speeds, deflections of such flaps provide differences in the pressures on upper and lower parts of the main surface ahead of flap, as well as on the flap itself At transonic speeds, the effect of the flap on the pressures on the main surface is greatly reduced because of the presence of local supersonic flows on this surface. As a result, the total forces produced by the flap are diminished. In addition, the hinge moments required to deflect the control may be greatly increased at transonic speeds

Corrective means. Various means are used to delay and reduce the adverse transonic characteristics. Among these are sweepback, that is, turning the wing or tail panels back to an oblique angle with the flight direction, reductions of the thickness-to-chord and span-to-chord (aspect) ratios for the wing or tail, additions on the wing; special fuselage contours; and rearrangement of the airplane components.

Sweepback The most effective means for improving the over-all transonic characteristics is to mount the wings slanting backward The action of such sweep may be most readily understood by considering the airflow over a very long swept surface (Fig. 3) Fundamentally, only the component of airflow normal to swept elements of this panel is effective in determining the nature of flow over the surface. Thus on such a swept surface, the onset of a shock wave, with the associated separation, is delayed until the reduced component of local velocity normal to the swept elements becomes supersonic. For substantial amounts of sweep, the flight speed may be high before this condition is reached. The use of sweep also greatly reduces the magnitude of the changes in the aerodynamic characteristics, once they occur. *See* WING

Theory and experiments indicate that the transonic characteristics are progressively delayed and reduced by increasing the sweep to relatively high values, but large amounts of wing sweep are not

sion and by changes in the source of power for propelling the bigger ships. A good example of recent projects to provide deeper inland shipping facilities is the St. Lawrence Seaway, completed jointly in 1959 by Canada and the United States to permit ocean-going ships of 27-ft draft to travel from the Atlantic Ocean as far west as Duluth, Minnesota.

At present oil-burning equipment provides power for most larger vessels, but a program has begun that is expected to produce many nuclear-powered vessels for commercial shipping. The NS (Nuclear Ship) *Savannah*, a large merchant ship was christened by the United States government in 1959. Other nations are engaged in building nuclear-driven ships. See REACTOR, SHIP PROPULSION.

Railway transportation. Engineering relating to rail transportation includes the planning and construction of terminals, switchyards, loading and unloading facilities, trackage, bridges, traffic-control and maintenance facilities, and the hauling equipment itself—locomotives and other rolling stock. The railroad industry has underway continuing programs to develop safer and quicker methods of loading, unloading, and shifting of cars, and of operating trains. See RAILROAD ENGINEERING.

Air transportation. The planning, design, and construction of runways, terminals, aircraft, and navigation aids are the major branches of civilian air transportation engineering. Width, length, strength, and layout of the runways are dependent upon the type of aircraft to be accommodated and the frequency of landings and departures. In like manner, design of the airport terminal building is determ...

press

terms ... to

helicopter traffic.

Recent introduction of jet-powered planes and continued emphasis on larger and faster aircraft are important factors in air transport engineering. See ...

pr

se

and ... water supply, sewage disposal

Pipelir

and cons

to move ... through the line and any needed storage facilities. Frequently large bridge structures are required to carry the pipeline across major streams or other barriers. In designing the pipeline consideration must be given to possible alternate routes, topography, right-of-way needs, foundation conditions, the desired pipe size, thickness of pipe, methods of laying, type of material to be transported, measures to protect the pipe once in place, forces on the conduit due to temperature changes and pumping operations, and adequate maintenance ... cedures.

Planning of pumping facilities requires study of power requirements for different types of material moved, standby facilities, spacing of the pumping stations, and related factors. See PIPELINE.

[A.S.C.]

Transuranium elements

Those elements of the actinide series from neptunium (atomic number 93) through element 102. As members of the actinide series they have a close chemical resemblance to the lanthanide, or rare-earth, series. Of these elements, only plutonium has been prepared on a large scale, while some of the others have been produced in barely weighable amounts. Neptunium is prepared, but not separated, as one step in the production of plutonium, which is of value in nuclear weapons and in the development of nuclear power for industrial purposes due to its fissionable property.

The concept of atomic weight in the sense applied to naturally occurring elements is not applicable to the transuranium elements, since the isotopic composition of any given sample depends on its source. In most cases the use of the mass number of the longest lived isotope in combination with an evaluation of its availability has been adequate. Good choices at present are neptunium, 237; plutonium, 242; americium, 243; curium, 248; berkelium, 249; californium, 249; einsteinium, 254; and fermium 255.

It appears that the eventual production of elements 103 through about 108 will be achieved by the bombardment of transuranium elements with heavy ions (heavier than helium ions). The half-lives of these elements are expected to be so short as to make conventional chemical identification difficult up to elements 104 and 105 and probably impossible beyond these. Possibly chemical identification can be made in some cases by using simple and fast methods, as for example, those involving migration of gaseous atoms or ions, volatility properties, reactions with surfaces, or gas flow reactions. It is likely that the present basic requirements for the discovery of a new element, namely complete chemical identification and separation from all previously known elements, will have to be changed at some point.

The actinide series should be completed at the undiscovered element 103, and the succeed...

... (iridium) to 76 (osmium). This analogy would continue in the still heavier transuranium elements, and element 118 would be a rare gas.

Brief descriptions of the transuranium elements follow. They are listed according to increasing atomic number. See also separate articles on each element.

Neptunium (Np, atomic number 93, after the planet Neptune). Neptunium was the first transuranium element discovered. In 1940, E. M. McMillan and P. H. Abelson at the University of

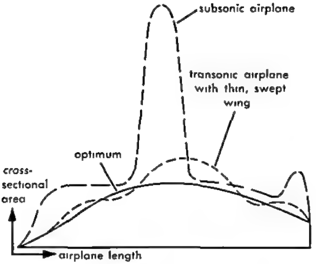


Fig. 5 Longitudinal area developments for various types of airplanes.

However, because of practical considerations, the lengths and cross-sectional areas of airplanes must be limited to values corresponding to a body of revolution with a ratio of length to diameter of about 9.

The longitudinal developments of area for conventional subsonic airplanes differ greatly from the ideal shape and as a result, the maximum transonic drag for such airplanes may be as much as ten times subsonic drag values. The various wing features used for delaying and reducing the transonic characteristics—sweep, thinner sections, and lower aspect ratios—usually result in airplane area developments more nearly approaching the most satisfactory shape (Fig 5) Consequently, the transonic drag for such airplanes is reduced to approximately three times the subsonic level

The area developments for some transonic and supersonic airplanes have been made to approach the form for lowest drag by special shaping of the fuselage. Such a shaping has been provided through the subtraction of fuselage volume in the region of the wing and tail, as well as by the addition of volume ahead of and behind these surfaces The transonic drag for airplanes with such shaped fuselages is as low as twice the subsonic values. The area development may also be improved by specially locating external bodies such as engine nacelles

The jet engines used to power transonic airplanes have been placed in various locations: within the fuselage or wing, in housings flush to these components, or on struts attached to these elements. Each location offers certain individual aerodynamic and structural advantages For transonic airplanes, intakes for air to the engine are usually of simple streamlined forms. *See* AIRPLANE; SUBSONIC FLIGHT; SUPERSONIC FLIGHT

[R.T.WH.]

Bibliography: A. F. Donovan and H R. Lawrence (eds), *Aerodynamic Components of Aircraft at High Speeds*, 1957; A. F. Donovan et al (eds.), *High Speed Problems of Aircraft and Experimental Methods*, 1960

Transparent medium

Ordinarily, a medium which has the property of transmitting rays of light in such a way that the human eye may see through the medium distinctly It is pervious to light, that is, to the visible region of the electromagnetic spectrum. By extension, a medium may be described as transparent to other regions of the spectrum such as x-rays and microwaves. Just as a blue filter passes blue rays, an ultraviolet filter might be considered as passing ultraviolet rays, or being transparent to them *See* ABSORPTION (ELECTROMAGNETIC RADIATION); TRANSLUCENT MEDIUM. [M G M]

Transportation engineering

That branch of engineering relating to the movements of goods and people. Major types of transportation are highway, water, rail, air, and pipeline.

Highway transportation. Highway transportation engineering deals with the planning, construction and operation of roads, streets, bridges, and parking facilities (*see* HIGHWAY ENGINEERING). Important aspects of highway engineering include:

1. Traffic engineering, which relates to the volumes of traffic to be handled, how best to accommodate these flows, and the general layout of highways *See* TRAFFIC ENGINEERING

2 Engineering pertaining to pavements and roadway surfaces *See* PAVEMENT.

3 Design and construction of highway bridges, structures, tunnels, and similar facilities *See* BRIDGE; TUNNEL.

In recent years highway transportation engineering has been distinguished by the development of planning and construction techniques that have made possible express highways to accommodate large flows of traffic at high speeds Express routes are now being built in the United States, Canada, Mexico, Europe, South America, and other areas Need for the express highways has resulted from the enormous growth of motor vehicle transportation For example, in the United States, where motor vehicle registration has grown from about 32,000,000 vehicles in 1940 to over 69,000,000 in 1959, the Federal government in cooperation with all state highway departments is building a 41,000-mile interstate highway system This network will cost over $10,000,000,000 and will consist entirely of freeways of high capacity.

Water transportation. Planning and construction of canals, channels, harbor facilities, navigation aids such as lighthouses, and navigation locks and dams are important concerns of water transportation engineers (*see* CANAL; COASTAL ENGINEERING; RIVER ENGINEERING). Another important field of water transportation relates to the design and production of launches, barges, tugs ferry boats, and other ships.

Water transportation engineering in recent years has been distinguished by the emphasis on larger equipment requiring increased waterway dimen

borg in late 1949 at the University of California in Berkeley and was the fifth transuranium element discovered. The isotope Bk^{243} was synthesized by helium-ion bombardment of Am^{241}. The first isolation of berkelium in weighable amount, as Bk^{249} (half-life 290 days), produced by neutron irradiation, was accomplished in 1958 by Thompson and Cunningham. Isotopes of mass numbers 243–250 are known

The chemical properties of berkelium have been studied with tracer amounts, and berkelium has been found to exist in the III and IV oxidation states in aqueous solution Its compounds appear to resemble those of the other actinide elements

Californium (Cf, atomic number 98, after the state and University of California) The sixth transuranium element to be discovered, californium in the form of the isotope Cf^{245} was prepared by the helium-ion bombardment of microgram quantities of Cm^{242} at the University of California at Berkeley The element was discovered by Thompson, K Street, Jr, Ghiorso, and Seaborg at the University of California in Berkeley early in 1950 Isotopes of mass numbers 244–254 are known The existence of the isotopes Cf^{249}, Cf^{250}, Cf^{251}, and Cf^{252} produced by neutron irradiation, with their relatively long half-lives, makes it feasible to isolate californium in weighable amount

Investigation of the chemical properties of californium has been performed with tracer amounts of the element and only the III oxidation state has been found to be stable in aqueous solution.

Einsteinium (Es, atomic number 99, after Albert Einstein) The seventh transuranium element to be discovered, einsteinium was found in the debris from the "Mike" thermonuclear explosion staged by the Los Alamos Scientific Laboratory in November 1952 Very heavy uranium isotopes were formed by the action of the intense neutron flux on the uranium in the device, and these decayed into isotopes of elements 99, 100, and other transuranium elements of lower atomic number. Chemical investigation of the debris in late 1952 by workers at the University of California Radiation Laboratory ...

... been synthesized The existence of the isomeric form of the isotope Es^{254} (half life 280 days) which can be prepared by neutron bombardment of Pu^{239} and other nuclides makes it feasible to produce the element in weighable amount.

The chemical properties of einsteinium have been studied on the tracer scale and the present knowledge of the element consists chiefly of its ion-exchange behavior Indications are that einsteinium in aqueous solution behaves as a tripositive actinide ion

Fermium (Fm, atomic number 100, after Enrico Fermi) Fermium, the eighth transuranium element discovered, was isolated as the isotope Fm^{255} from the heavy elements formed in the "Mike" thermonuclear explosion. The element was discovered in early 1953 by Ghiorso and coworkers during the same investigation which resulted in the discovery of element 99. Fermium isotopes of mass number 248–256 have been prepared Since Fm^{255}, the longest lived fermium isotope known, has a half-life of 22 hours, it is improbable that weighable amounts of fermium will ever be obtained.

Studies of the chemical properties of fermium, performed on the tracer scale, indicate that fermium in aqueous solution behaves as a typical tripositive actinide ion.

Mendelevium (Md, atomic number 101, after Dmitri Mendeleev). Mendelevium, the ninth transuranium element discovered, was identified by Ghiorso, B G, Harvey, G. R. Choppin, Thompson, and Seaborg at the University of California in Berkeley in 1955. The element as Md^{256} was produced by the bombardment of extremely small amounts (approximately 10^9 atoms) of Es^{253} with helium ions The first identification of einsteinium was notable in that only one or two atoms per experiment were produced. There appear to be no isotopes of sufficiently long half-life to make possible the isolation of mendelevium in weighable amount.

The chemical properties have been investigated on the tracer scale and the element found to behave in aqueous solution as a typical tripositive actinide ion

Nobelium. The isotope No^{254} with a half-life of about 3 sec was discovered by Ghiorso, T. Sikkeland, J R. Walton, and Seaborg at the University of California in Berkeley in 1958 The element was produced by the bombardment of Cm^{246} with C^{12} ions accelerated in the heavy ion linear accelerator The isotope was identified through chemical identification of the known daughter, Fm^{250}, which was separated by recoil from its alpha-decaying parent See ACTINIDE ELEMENTS; NUCLEAR CHEMISTRY; NUCLEAR REACTION, PERIODIC TABLE. [G T S.]

Bibliography: S Glasstone, *Sourcebook on Atomic Energy*, 2d ed, 1958; J. J. Katz and G. T. Seaborg, *The Chemistry of the Actinide Elements*, 1957; G. T Seaborg, *The Transuranium Elements*, 1958, D Strominger, J M Hollander, and G. T Seaborg, Table of isotopes, *Revs Modern Phys.*, 1958.

Trapezoid

A term used in the United States for a quadrilateral with two sides parallel In Great Britain such a figure is often called a trapezium, a term used in the United States for a general quadrilateral. The parallel sides are called the bases, and the area of a trapezoid is the product of one-half the sum of the bases by the perpendicular distance between them (altitude) If the two nonparallel sides (legs) are equal, the trapezoid is isosceles. A procedure for approximating the value of a definite integral

$$\int_a^b f(x)\,dx$$

California in Berkeley identified the isotope Np^{239} which was produced by the bombardment of uranium with neutrons according to the reaction

$$U^{238}(n,\gamma)U^{239} \rightarrow Np^{239}$$

The element as Np^{237} was first isolated as a pure compound, the oxide, in 1944 by L. B. Magnusson and T. J. La Chapelle. Neptunium in trace amounts is found in nature, the element being produced in nuclear reactions in uranium ores caused by the neutrons present. Gram and larger quantities of Np^{237} (half-life 2.2×10^6 years) are being produced as a by-product of the production of plutonium in nuclear reactors. Isotopes from mass number 231 to 241 have been synthesized.

Neptunium displays four oxidation states in aqueous solution: Np^{3+} (blue to purple), Np^{4+} (yellow-green), NpO_2^{+} (green), and NpO_2^{++} (pink to red). The ion NpO_2^{+}, unlike corresponding ions of other actinide elements, can exist in aqueous solution at moderately high concentrations

Neptunium metal has a silvery appearance, is chemically reactive, and melts at 640°C; the solid metal has at least three crystalline forms between room temperature and its melting point

Plutonium (Pu, atomic number 94, after the planet Pluto). Plutonium in the form of Pu^{238} was discovered in late 1940 by G. T Seaborg, E M. McMillan, J. W. Kennedy, and A. C. Wahl at the University of California in Berkeley. The element was produced in the bombardment of uranium with deuterons according to the reaction

$$U^{238}(d,2n)\ Np^{238} \xrightarrow[2\ 1\ \text{days}]{\beta^-} Pu^{238}$$

Plutonium as Pu^{239} was first isolated as a pure compound, the fluoride, in 1942 by B B. Cunningham and L. B. Werner. Minute amounts of plutonium formed in much the same way as naturally occurring neptunium are present in nature. The important isotope Pu^{239} was discovered by Kennedy, Seaborg, E Segrè, and Wahl in 1941 Plutonium-239 (half-life 24,360 years), owing to its property of being fissionable with neutrons, has been used successfully as the explosive ingredient in nuclear weapons and is a key material in the development of nuclear energy for industrial purposes, 1 pound being equivalent to about 10,000,000 kilowatthours of heat energy. Methods have been developed for the production of Pu^{239} in kilogram and larger amounts from natural uranium in nuclear reactors The alpha radioactivity and physiological behavior of this isotope make it one of the most dangerous poisons known. Isotopes of mass number 232–246 are known. The isotopes Pu^{242} and (eventually) Pu^{244} are more suitable than Pu^{239} for chemical investigation because of their longer half-lives and lower specific activities

Plutonium has four oxidation states in aqueous solution: Pu^{3+} (blue to violet), Pu^{4+} (tan to orange-brown), PuO_2^{+} (predicted to be reddish-purple), and PuO_2^{++} (yellow to pink-orange). The ions Pu^{4+} and PuO_2^{+} undergo extensive disproportionation to the ions of higher and lower oxidation state. All four oxidation states can exist simultaneously at appreciable concentrations in equilibrium with each other, an unusual situation that leads to complicated solution phenomena

The metal is silvery in appearance, is chemically reactive, melts at 639.5°C, and has six crystalline modifications between room temperature and its melting point

Americium (Am, atomic number 95, after the Americas). Americium was the fourth transuranium element discovered. The element as Am^{241} (half-life 458 years) was produced by the intense neutron bombardment of plutonium and was identified by Seaborg, R. A. James, L O. Morgan, and A. Ghiorso in late 1944 and early 1945 at the wartime Metallurgical Laboratory at the University of Chicago. Using the isotope Am^{241}, the element was first isolated as a pure compound, the hydroxide, in 1945 by B B Cunningham. Isotopes of mass numbers 237–246 have been prepared Gram and larger quantities of Am^{241} are being produced. The less radioactive isotope Am^{243} is more suitable for use in chemical investigation.

Americium exists in three oxidation states in aqueous solution: Am^{3+} (pink), AmO_2^{+} (yellow), and AmO_2^{++} (rum-colored). The ion AmO_2^{+} is unstable with respect to disproportionation into Am^{3+} and AmO_2^{++}. Americium displays the IV state in solid compounds only.

Metallic americium is silvery-white in appearance, and its melting point is probably below 900°C. It is more electropositive than plutonium, being comparable to the lighter rare-earth metals in this respect Americium metal has at least three crystalline modifications between room temperature and its melting point

Curium (Cm, atomic number 96, after Pierre and Marie Curie) The third transuranium element to be discovered, curium as the isotope Cm^{242} was identified by Seaborg, James, and Ghiorso in 1944 at the wartime Metallurgical Laboratory of the University of Chicago Curium was produced by the helium-ion bombardment of Pu^{239} in the University of California 60-in cyclotron Curium was first isolated, using the isotope Cm^{242}, in the form of a pure compound, the hydroxide, in 1947 by L. B. Werner and I. Perlman. Isotopes of mass number 238–250 are known Chemical investigations with macroscopic quantities of curium have been performed, using Cm^{242} (half-life 162.5 days) and Cm^{244} (half-life 19 years), Cm^{244} and higher mass isotopes are more satisfactory for this purpose; these are all produced by neutron irradiation.

The only oxidation state of curium in aqueous solution appears to be the III state The metal is silvery and shiny in appearance, and indications are that it is somewhat more reactive than the lighter actinide metals

Berkelium (Bk, atomic number 97, after Berkeley, California). Berkelium was produced and identified by S. G. Thompson, Ghi and Sea-

borg in late 1949 at the University of California in Berkeley and was the fifth *transuranium* element discovered. The isotope Bk^{243} was synthesized by helium ion bombardment of Am^{241}. The first isolation of berkelium in weighable amount, as Bk^{249} (half-life 290 days), produced by neutron irradiation was accomplished in 1958 by Thompson and Cunningham. Isotopes of mass numbers 243–250 are known

The chemical properties of berkelium have been studied with tracer amounts, and berkelium has been found to exist in the III and IV oxidation states in [illegible]

to rese [illegible]

Calif [illegible]

state and University of California). The sixth transuranium element to be discovered, californium in the form of the isotope Cf^{245} was prepared by the helium ion bombardment of microgram quantities of Cm^{242} at the University of California at Berkeley. The element was discovered by Thompson, K. Street, Jr., Ghiorso, and Seaborg at the University of California in Berkeley early in 1950. Isotopes of mass numbers 244–254 are known. The existence of the isotopes Cf^{249}, Cf^{250}, Cf^{251}, and Cf^{252} produced by neutron irradiation, with their relatively long half-lives, makes it feasible to isolate californium in weighable amount

Investigation of the chemical properties of californium has been performed with tracer amounts of the element, and only the III oxidation state has been found to be stable in aqueous solution.

Einsteinium (Es, atomic number 99, after Albert Einstein). The seventh transuranium element to be discovered, einsteinium was found in the debris from the "Mike" thermonuclear explosion staged by the Los Alamos Scientific Laboratory in November 1952. Very heavy uranium isotopes were formed by the action of the intense neutron flux on the uranium in the device, and these decayed into isotopes of elements 99, 100, and other transuranium elements of lower atomic number. Chemical investigation of the debris in late 1952 by workers at the University of California Radiation Laboratory, Argonne National Laboratory, and Los Alamos Scientific Laboratory revealed the presence of element 99 as the isotope Es^{253}. Isotopes of mass numbers 246–256 have been synthesized. The existence of the isomeric form of the isotope Es^{254} (half-life 280 days), which can be prepared by neutron bombardment of Pu^{239} and other nuclides makes it feasible to *produce* the element in weighable amount

The chemical properties of einsteinium have been studied on the tracer scale, and the *present* knowledge of the element consists chiefly of its ion-exchange behavior. Indications are that einsteinium in aqueous solution behaves as a tripositive actinide ion

Fermium (Fm, atomic number 100, after Enrico Fermi). Fermium, the eighth transuranium element discovered was isolated as the isotope Fm^{255} from the heavy elements formed in the "Mike" thermonuclear explosion. The element was discovered in early 1953 by Ghiorso and coworkers during the same investigation which resulted in the discovery of element 99. Fermium isotopes of mass number 248–256 have been prepared. Since Fm^{253}, the longest-lived fermium isotope known, has a half life of 22 hours, it is *improbable that* weighable amounts of fermium will ever be obtained.

Studies of the chemical properties of fermium, performed on the tracer scale, indicate that fermium in aqueous solution behaves as a typical tripositive actinide ion.

Mendelevium (Md, atomic number 101, after Dmitri Mendeleev). Mendelevium, *the ninth transuranium element discovered*, was *identified* by Ghiorso, B. G. Harvey, G. R. Choppin, Thompson, and Seaborg at the University of California in Berkeley in 1955. The element as Md^{256} was produced by the bombardment of extremely small amounts (approximately 10^9 atoms) of Es^{253} with helium ions. The first identification of einsteinium was notable in that only one or two atoms per experiment were produced. There appear to be no isotopes of sufficiently long half-life to make possible the isolation of mendelevium in weighable amount.

The chemical properties have been investigated on the tracer scale and the element found to behave in aqueous solution as a typical tripositive actinide ion

Nobelium. The isotope No^{254} with a half-life of about 3 sec was discovered by Ghiorso, T. Sikkeland, J. R. Walton, and Seaborg at the University of California in Berkeley in 1958. The element was produced by the bombardment of Cm^{246} with C^{12} ions accelerated in the heavy ion linear accelerator. The isotope was identified through chemical identification of the known daughter, Fm^{250}, which was separated by recoil from its alpha-decaying parent. *See* ACTINIDE ELEMENTS, NUCLEAR CHEMISTRY, NUCLEAR REACTION, PERIODIC TABLE [C.T.S.]

Bibliography: S. Glasstone, *Sourcebook on Atomic Energy*, 2d ed., 1958; J. J. Katz and G. T. Seaborg, *The Chemistry of the Actinide Elements*, 1957; G. T. Seaborg, *The Transuranium Elements*, 1958; D. Strominger, J. M. Hollander, and G. T. Seaborg, Table of isotopes, *Revs. Modern Phys.*, 1958.

Trapezoid

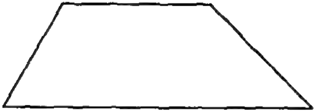
A term used in the United States for a quadrilateral with two sides parallel. In Great Britain such a figure is often called a trapezium, a term used in the United States for a general quadrilateral. The parallel sides are called the bases, and the area of a trapezoid is the product of *one-half* the sum of the bases by the [illegible]

them (altit [illegible]

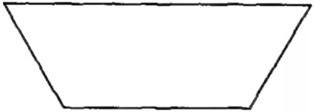
are equal, [illegible]

for approxi [illegible]

$$\int_a^b f(x)\,dx$$



(a)



(b)

(a) Trapezoid. (b) Isosceles trapezoid

consists of adding the areas of trapezoids formed by dropping perpendiculars from points of the curve $y = f(x)$, $a \leqq x \leqq b$, to the x axis It is known as the trapezoidal rule *See* INTEGRATION; QUADRILATERAL. [L M.BL.]

Traps in solids

Electron traps are defects or chemical impurities in semiconductors and in insulators which capture mobile electrons in a special way The electrons are immobilized, prevented from recombining with (annihilating) positive holes, and are released some time later as mobile electrons again Analogous traps act on mobile positive holes A defect which captures a mobile electron and aids recombination with a hole (or vice versa) is called a recombination center A defect (nickel in germanium, for example) may be a recombination center at one temperature and a trap at a different temperature.

Electrons or holes may remain in traps for short times, or for as long as months or years They may be released by heating the solid or by irradiating it with infrared. Traps have an important influence in photoconduction, luminescence, and the photographic process. *See* BAND THEORY OF SOLIDS, HOLES IN SOLIDS; LUMINESCENCE; PHOTOCONDUCTIVITY; PHOTOLYSIS (PHOTOCHEMISTRY) [L A]

Trauma

Injury to tissue by physical or chemical means Mechanical injury includes abrasions, contusions, lacerations, and incisions, as well as stab, puncture, and bullet wounds. Trauma to bones and joints results in fractures, dislocations, and sprains Head injuries are often serious because of the complications of hemorrhage, skull fracture, or concussion.

Thermal, electrical, and chemical burns produce severe damage partly because they coagulate tissue and seal off restorative blood flow. Asphyxiation including that caused by drowning, produces rapid damage to the brain and respiratory centers, as well as to other organs.

Frequent complications of trauma are shock, the state of collapse precipitated by peripheral circulatory failure, and also hemorrhage, infection, and improper healing [E.G.ST]

Traveling-wave tube

A microwave electron tube in which the beam of electrons interacts continuously with a wave traveling on a circuit over appreciable distance. Its normal use is as an amplifier with an exceedingly wide bandwidth as compared with triodes, klystrons, or other amplifiers utilizing tuned circuits to produce gain. Bandwidths of from 10–100% of the center frequency are common, with gains from 20–60 decibels Traveling-wave tubes for the input to sensitive radar or communication receivers have been built with noise figures as low as 3.5 decibels. At the other extreme, high-power traveling-wave tubes for the final amplifier stages of radar or scatter communication systems have been built with pulsed powers over a megawatt.

A schematic diagram of a traveling-wave tube is shown in Fig. 1 The beam of electrons is produced by a thermionic cathode and focused by an electron gun and some additional focusing method through out the length of the beam. The radio-frequency energy is introduced on a slow-wave circuit, producing an electromagnetic wave on the circuit which propagates down the tube at approximately the same velocity as that of the electron beam Interaction between beam and circuit is then continuous with contributions adding in phase, and power is taken by a matched load at the output

Electron gun. The electron gun for a traveling-wave tube is usually of the Pierce type (Fig 2*a*) in which the plane face of the cylinder is the active emitting surface and is followed by a focusing electrode and an anode The beam leaves the gun straight and parallel in preparation for entrance into the interaction region Since the electron flow is nearly parallel, current density in the beam is limited to that available from the oxide cathode (a fraction of an ampere per square centimeter) The

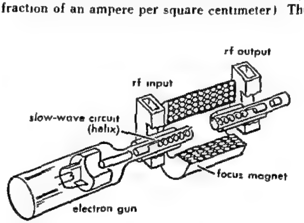


Fig 1. Schematic diagram of traveling-wave tube

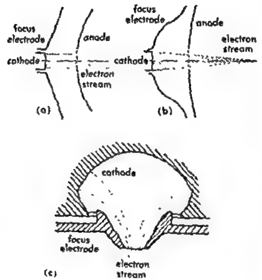


Fig 2 Electron guns for beam-type tubes of the traveling-wave class. (a) Pierce gun (b) Pierce gun with converging flow (c) Heil gun

perveance of such guns is difficult to make much over 1×10^{-6} where

$$\text{Perveance} = \frac{\text{dc current}}{(\text{dc beam voltage})^{3/2}} \qquad (1)$$

For higher-perveance guns and particularly where higher current densities are required than those obtainable directly from the cathode surface, converging flow guns are designed Some are based upon

... may be used.

For higher-perveance guns in high power tubes or backward wave tubes, it may be desirable to use hollow electron beams The Pierce design for a hollow beam is similar to that for the solid beam ex-

...beams can also be made (although somewhat more difficult to design than those for solid beams) using a similar principle, as illustrated in Fig 3b

For low-noise guns, additional noise reducing electrodes are introduced to modify the space-charge waves of noise produced by the statistical fluctuation of current and velocity at the cathode

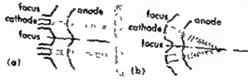


Fig 3 Hollow beam electron guns. (a) Parallel flow. (b) Converging flow

This principle was discovered by D. A. Watkins. An early gun used a series of cylinders to produce jumps in velocity for noise reduction as in Fig. 4a Present low-noise guns simply have a series of electrodes on which different dc potentials may be placed (Fig. 4b) to produce the dc potential profile along the beam for optimum noise reduction. The low noise figures referred to in the first paragraph have been obtained in this way, whereas a tube without such noise reduction would have noise figures from 15–30 decibels

Focusing methods for the beam. To keep the beam focused throughout the rf interaction region, some method of beam focusing is normally required to overcome the space-charge forces of the relatively high-density beams utilized in traveling-wave tubes. The simplest method is that of providing a longitudinal field throughout the tube (Fig. 5a) and is called confined flow By this, any tendency for radial expansion is converted to a tight spiralling motion of the electrons by the magnetic forces. The magnetic field required is typically from a few hundred to a few thousand gauss and may be obtained either by electromagnets or permanent magnets. For practical applications, the magnets may be quite heavy, and the power required by an electromagnet may also be undesirable

A similar method of focusing, called space-charge balanced flow, utilizes a lower value of magnetic field in the gun region than in the interaction region Because of this, the entire beam is given a rotation at the transition between the two values of axial field, and centrifugal forces arising from this rotation play a part in the focusing. The special case of this, with no magnetic field at the cathode, called Brillouin flow, is illustrated in Fig 5b, and is required for use with convergent-beam guns where no magnetic forces are required in the converging region Harris flow (Fig. 5c), used with hollow beams, also produces a beam rotation by a discontinuity in axial magnetic field values However, here a radial dc electric field is added to

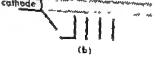


Fig 4 Low-noise guns for traveling-wave tubes. (a) Watkins type with velocity jumps (b) Gradual acceleration type

produce electrical forces as well as the centrifugal and space-charge forces.

Systems of magnetic or electrostatic lenses for focusing have been developed for many traveling-wave tubes and produce a much lighter focusing system. The system of magnetic lenses requires a periodic magnetic field, usually produced by permanent magnets, as shown in Fig. 5*d*. If ferrites are used, the lens system can be light in weight. The system of electrostatic lenses is potentially attractive, because it eliminates even the small periodic magnets. This system has been limited in application, because it is easy to use only with circuits that have natural separations for application of the different dc voltages. The folded line circuit of Fig. 5*e* can have a different dc potential on the top line from that on the bottom, producing a periodic electrostatic field and an electron-lens system.

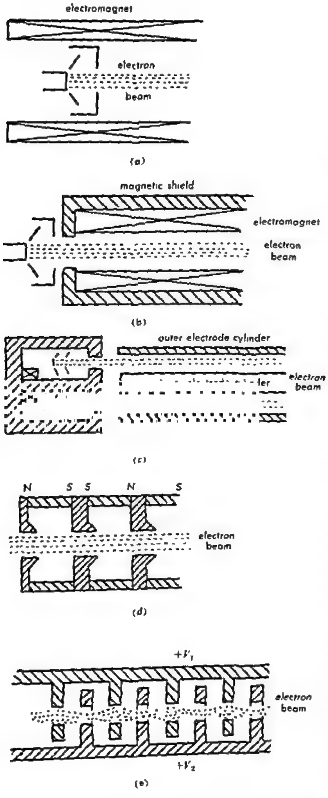


Fig. 5 Beam-confining methods (*a*) Confined flow. (*b*) Brillouin flow. (*c*) Harris flow. (*d*) Periodic magnetic focusing (*e*) Electrostatic focusing.

Slow-wave circuit. The slow-wave circuit of a traveling-wave tube is designed to produce an electromagnetic wave of velocity approximately equal to that of the beam, which is typically from 1/50 to 1/10 the velocity of light. The wave velocity must remain substantially constant over the *bandwidth* of operation desired for the tube. The field produced by the circuit at the beam for a given power on the circuit, represented by the *interaction impedance*

$$K = \frac{(\text{ac electric field at beam})^2}{2(\text{average power flow on circuit})\beta^2} \qquad (2)$$

is another important characteristic; β is the *phase constant*.

A helix as pictured in Fig. 1 and Fig. 6*a* is one of the simplest and best slow-wave circuits and is used in a majority of traveling-wave tubes. The electromagnetic wave travels along the wire at approximately the velocity of light, so that its phase velocity along the axial direction is

$$v_p = \text{velocity of light}\left(\frac{\text{pitch}}{\text{circumference}}\right) \qquad (3)$$

In this approximation, phase velocity is independent of frequency, making possible wide-band amplifiers. The helix also has good interaction impedance.

For high-power tubes, the helix may be water cooled, but even then it may have difficulty in dissipating the heat losses. Other circuits of the loaded or folded waveguide type (Figs. 6*b* and 5*e* respectively) may have better power handling capabilities than the helix, although the former usually have smaller bandwidths. A 10% bandwidth may be typical for a high-power tube.

All of the circuits pictured are periodic and so have an infinite number of space harmonics. Some of these higher-order harmonics may be used instead of the fundamental for traveling-wave tubes. A good circuit developed by A. Karp for the very high frequencies (millimeter wavelength) is shown in Fig. 6*c*.

Other problems concerned with the circuit are to provide coupling in and out of the circuit and to provide attenuation on the circuit so that reflected waves do not cause the tube to oscillate. The couplings must be matched to eliminate reflections over the desired operating band. The attenuator problem is relatively easy for low-power tubes, being taken care of by glossy material sprayed on the circuit or

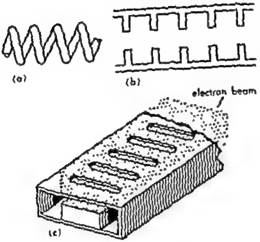
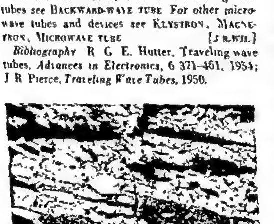


Fig 6 Typical slow wave structures for traveling-wave tubes (a) Helix (b) Loaded wave guide (c) Karp structure

its dielectric supporting structure, but may be difficult for high-power tubes because of the dissipation in the attenuator Unilateral attenuators using ferrites are ideal because they introduce no loss in the desired direction of propagation, but these have been successfully combined with slow-wave structures in only a few cases

For the backward wave class of traveling-wave tubes *see* BACKWARD-WAVE TUBE For other microwave tubes and devices *see* KLYSTRON, MAGNETRON, MICROWAVE TUBE [J R.WH.]

Bibliography R G E. Hutter, Traveling wave tubes, *Advances in Electronics*, 6 371–461, 1954; J R Pierce, *Traveling Wave Tubes*, 1950.



Travertine, Suisun, California Nicols crossed (From E Wm. Heinrich, *Microscopic Petrography*, McGraw-Hill, 1956)

Travertine

A rather dense, banded limestone, sometimes moderately porous, that is formed either by evaporation about springs, as is tufa, or in caves, as stalactites, stalagmites, or dripstone Where travertine or tufa (calcareous sinter) is deposited by hot springs, it may be the result of the loss of carbon dioxide from the waters as pressure is released upon emerging at the surface; the release of carbon dioxide lowers the solubility of calcium carbonate and it precipitates. High rates of evaporation in hot-spring pools also lead to supersaturation. Travertine formed in caves is simply the result of com- [illegible] ainly cal- [illegible] ITES AND [illegible] [R.S.]

Tree

A tree may be defined as a woody plant usually having one principal stem, or trunk, with a definite crown of branches and leaves. Some authorities have maintained that the trunk must be of a certain height and thickness to be considered a tree. However it is the form and not the size that constitutes the most widely accepted criterion. Even forms as small as the bonsai (Japanese dwarf trees) can properly be called trees

The map shows the position of the world's major forest regions (Fig 1) Distribution is governed primarily by climate, which is determined by elevation, by the relative position of adjacent high mountains and ocean currents, and by latitude. *See* VEGETATION ZONES (WORLD).

Each tree has a botanical or scientific name based on its taxonomic relationships to other plants, and one or more common names (*see* PLANT TAXONOMY) See separate articles on important trees listed by common names [A H G.]

Diseases of forest trees. From seed to maturity, forest trees are subject to a succession of diseases which annually may destroy timber equal to 45% of the saw timber growth Seedlings, especially conifers, are killed by soil inhabiting fungi (damping off), older seedlings may be attacked by an unknown complex of fungi and nematodes (*see* FUNGI, NEMATODA) Chemical treatment of soil with acids, fungicides, and biocides as well as cultural practices that are unfavorable to fungus growth, will help control seedling diseases (*see* FOREST ECOLOGY; FOREST SEEDING AND PLANTING, FUNGISTAT AND FUNGICIDE).

Usually in the forest, leaf diseases cause negligible losses, but brown spot needle blight (*Scirrhia acicola*) can cause excessive defoliation in nurseries and plantations and prevent longleaf pine from starting height growth (*see* FOREST TREE NURSERY MANAGEMENT) Fungicides in the nurseries and prescribed burning in plantations have been used successfully for control.

Diseases such as oak wilt (Fig 2) are systemic (infect entire plant body), killing susceptible species in a matter of weeks or months by plugging

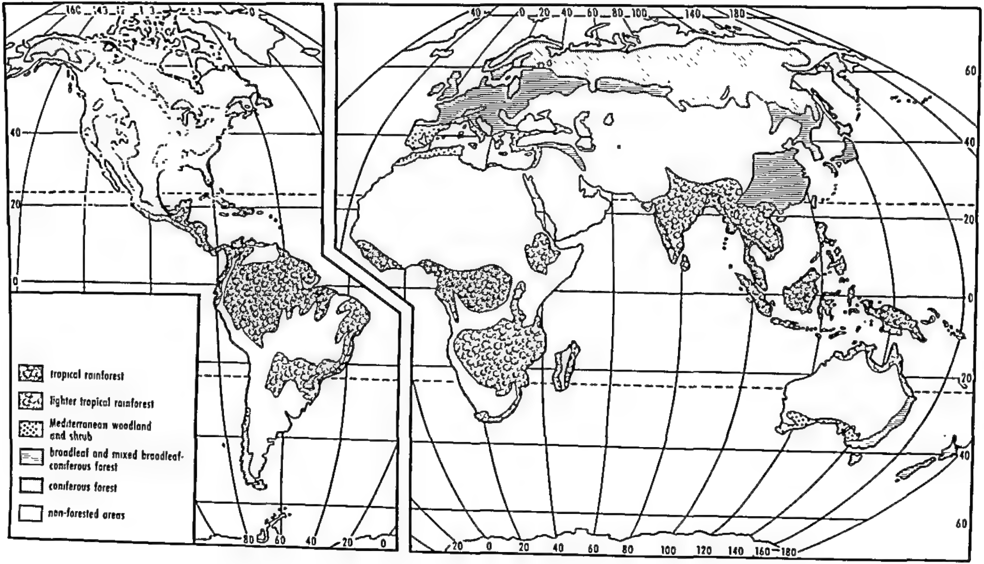


Fig 1. Major forest regions of the world.

the tree's vascular system (*see* VASCULAR BUNDLES). The fungus, *Ceratocystis fagacearum*, is disseminated to nearby healthy trees through root grafts and longer distances by the Nitidulidae and other insects (*see* COLEOPTERA; GRAFTING OF PLANTS). Currently, control is being attempted by eradication of infected trees.

Canker diseases range from chestnut blight, which has practically annihilated the American chestnut (Fig. 3), to less important cankers caused by species of *Nectria* and *Eutypella* which may only deform a portion of a tree. Satisfactory resistant varieties of chestnut may be developed, but for many canker diseases the *only control* measures are to remove and destroy infected trees in improvement cuttings.

The rust fungi, such as white pine blister rust, *Cronartium ribicola*, which attack the main stem of a tree are destructive (Fig. 4), but other rusts which invade only the needlelike leaves are not so serious. Usually each rust fungus has an unrelated alternate host, such as the wild currant for white pine blister rust. Eradication of the alternate host, resistant varieties, and selection of planting sites with less favorable environment for the fungus are major control measures.

All tree species, even woods resistant to decay such as redwood, are subject to one or more fungi causing heartrot (Fig. 5). Some of these fungi enter through wounds and branch stubs and decay the heartwood portion of the tree; others enter through roots and decay roots and the lower portion of the main stem. Shortening rotation age and avoiding wounds are methods of limiting heartrot losses.

Fig. 2 Oak wilt. (a) Oaks killed by the oak wilt fungus. (b) Mycelial *mat* that the oak wilt fungus produces beneath the bark of a wilted tree; the mycelial mats are covered with spores that can be disseminated by insects.

Fig. 3 Chestnut blight. (a) Blighted American chestnut tree. Large living native *chestnut* trees are rarely seen in northern states, and they are vanishing rapidly in the Smoky *Mountains* of the southern Appalachians (USDA). (b) *Mycelial* fan of chestnut blight, *Endothia parasitica*, advancing through bark of American chestnut. The tip of the fan on the left is *surrounded* by *cortical* tissue. The contents of the cortical cells back from the tip of the fan are discolored to a yellowish brown, as indicated by the darkened cells (after W. C. Bramble, from J. S. Boyce, *Forest Pathology*, 2d ed., McGraw-Hill, 1948)

Dwarf mistletoe is a small, partially parasitic plant that causes witches brooms (abnormal development of groups of small branches) and eventually kills ponderosa pine, other western conifers, and black spruce (Fig 6). Control consists of eradicating infected trees.

In the 1930s birch dieback, littleleaf disease of shortleaf pine, and pole blight of western white pine appeared, and as yet the causes are not completely known. Birch dieback may be due in part to increased soil temperature which causes excessive root mortality (*see* PLANT GROWTH). Littleleaf evidently results from a combination of poor soil drainage and a root parasite, *Phytophthora cinnamomi*. Pole blight may be due to soil moisture deficiencies (*see* PLANT, WATER RELATIONS OF). These exemplify diseases caused by unfavorable environment which may or may not involve a parasite.

Diseases of shade trees. Shade trees are often grown in abnormal habitats. This fact makes disease problems of shade trees somewhat different from those of forest trees, although many diseases affect both.

The desire for new and different shade trees has resulted in the introduction of exotic species, many of which are unsuited to the climate. Even native species are often taken from their normal habitats where they are protected by other trees and ground cover and placed in a more exposed and less favorable environment. Such trees require special care, including fertilizing and watering.

New disease problems often develop on introduced trees Colorado blue spruce in its natural habitat is seldom attacked by the spruce canker fungus, *Cytospora kunzei*, yet this fungus can kill blue spruce planted in other parts of the country.

Fig 5 Shelf fungus, *Fomes applanatus*, on a dead aspen tree. A wood rotting fungus that enters through roots and wounds, and decays the heartwood and sapwood of both hardwoods and softwoods (angiosperms and gymnosperms)

Fig 4. White pine blister rust, *Cronartium ribicola* (Photograph by Robert Campbell)

Fig 6 Pistillate shoots of ponderosa pine mistletoe (*Arceuthobium vaginatum forma cryptopodum*) on branch of sapling. Note unripe fruits (U S Forest Serv.)

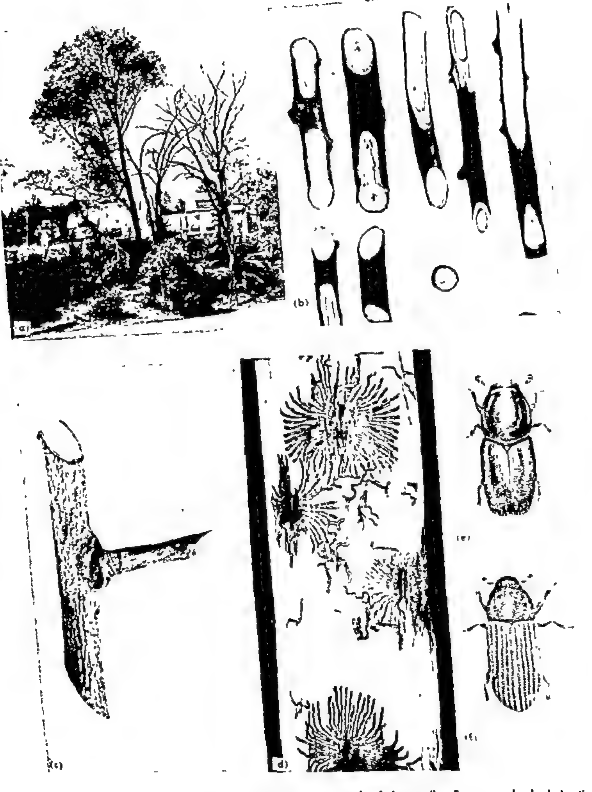


Fig 7 Dutch elm disease (a) Group of trees affected by the Dutch elm disease Elm bark beetles that had emerged earlier from the log in the foreground may have been instrumental in carrying the disease to the trees (USDA) (b) Discoloration in the sapwood of trees infected with Dutch elm disease (Michigan State University) (c) Feeding scar in small elm crotch made by an adult of the smaller European elm bark beetle, *Scolytus multistriatus* (USDA) (d) Brood galleries made by female beetles and larvae (USDA) (e) European elm bark beetle, the most important carrier of the Dutch elm disease (Michigan State University) (f) Native elm bark beetle (Michigan State University).

Fig. 8 Elms dying of phloem necrosis. (*USDA*)

Fig 9 Elm infected with bacterial wetwood. The bacterial ooze is coming from a small wound.

The effects of this secondary parasite can be minimized by selecting protected sites with better soils and avoiding dry southwest exposures As a general rule many physiological diseases, as well as those caused by weakly parasitic fungi, can be avoided by selecting native tree species grown from a local seed source.

Dutch elm disease, introduced in the United States in 1930 or earlier, is an important shade tree disease because of the value of the host (Fig. 7). The Dutch elm disease fungus, *Ceratocystis ulmi*, is introduced to the vascular system of healthy trees by the small European elm bark beetle, *Scolytus multistriatus*, and the native species, *Hylurgopinus rufipes*. Possibly resistant varieties can be developed, but until then controlling the insect vector, including the removal of dead and dying elms, is the best approach.

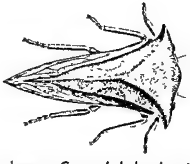
Elm is also attacked by a virus disease called phloem necrosis (Fig. 8), the only known important virus disease of shade trees (*see* PLANT VIRUS). Infected trees wilt and die, the inner bark turning brown and emitting a wintergreen odor. The virus is transmitted by the elm leafhopper, *Scaphoides luteolus* (*see* HOMOPTERA). One of the few known bacterial diseases of trees is wetwood of elm, caused by *Erwinia nimipressuralis* (Fig. 9). Wetwood in many other tree species is caused by an unknown complex of bacteria (*see* BACTERIA). Satisfactory control measures have not been found

Diseases of minor importance in the forest, such as leaf spots, may be very objectionable on shade trees. Many of these can be controlled by fungicides applied as the leaves unfold in the spring. On the other hand, heartrot is a major problem in the forest, but it is of minor importance in shade trees because such trees may remain alive for decades with extensive heartrot. Expensive cavity work is of questionable value, since it does not stop the decay process and may not add much strength to the tree. [D.W.FR]

Bibliography: See FOREST AND FORESTRY; PLANT DISEASE.

Treehopper

Any member of the insect family Membracidae, order Homoptera, Treehoppers are usually less than ½ in. long, and are frequently grotesque in shape due to the overdevelopment of the pronotum, the dorsal plate of the first thoracic segment. Although usually three-cornered, the pronotum on some species is prolonged into a thornlike structure Only a few of the 185 species in the United States are of economic importance. Best known is the buffalo treehopper, *Ceresa bubalus*, which kills the twigs of fruit trees by depositing eggs in them; and the three-cornered alfalfa hopper, *Spissistilus festinus*, which may become a pest by girdling the plant stems in feeding *See* HOMOPTERA [J.D B.]



Buffalo treehopper, *Ceresa bubalus;* length under ½ in. (*From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

Tree-ring hydrology

The relationships between ring thicknesses and stream flow, and between thicknesses and rainfall via soil moisture These relationships have validity only if it is possible to identify accurately the annual increments. *See* DENDROCHRONOLOGY.

In the western part of the United States ring thicknesses appear to have only a modest and rather inconsistent correlation with stream flow. The same may be said for the comparison of rainfall with tree growth of the lower forest border in the Southwest where agreement averages 40–70%. In the forest interior of northern New Mexico agreement between the variations of rainfall and tree growth exceeds 96%.

The use of growth ring patterns rather than individual rings eliminates the necessity of dating accurately each growth layer. An analysis of ring patterns and rainfall regimes gives three types in the Southwest and West. The California, or Sierra, type has a winter rainy season, and trees depend upon soil moisture augmented chiefly by rains before the start of growth. Hence, the pattern consists of growth layers rather uniform in thickness [illegible] the pattern therefore consists of growth layers highly variable in thickness, and multiple and partial within the annual increment The Arizona type has both a winter and summer rainy season Hence the [illegible] within the annual increment The West Texas type apparently extends westward at low elevations and the California type eastward at high elevations In the area of contact a tension zone exists in the forest; the contact rises during dry years and descends during wet ones

Thus growth patterns reveal information about rainfall through the variability in thickness of successive growth layers, relative and absolute variations in circuit and longitudinal uniformity, areal distribution of partial growth layers, and the constitution of the annual increment [W S C.]

Bibliography C G. Dobbs, A study of growth rings in trees *Forestry*, vol. 24 1951; W S Glock, Growth rings and climate, *Botan Rev*, vol 7, 1941 and vol 21 1955; W S Glock, Tree growth and rainfall *Trans Am. Geophys Union*, vol 36, 1955; E Huntington, *Tree Growth and Climatic Interpretations* Carnegie Inst Wash Publ 352 1925.

Trematoda

A class of acoelomate, parasitic flatworms of the phylum Platyhelminthes They commonly occur as adults in or on all vertebrate groups They exhibit cephalization bilateral symmetry, and well developed anterior and ventral, or anterior and posterior, holdfast structures. A cuticle is present in place of a true epidermis. The mouth is anterior and, usually, a blind, forked gut occurs, as well as three muscle layers. The excretory system consists of flame cells and collecting tubules. These animals are predominantly hermaphroditic and oviparous with operculated egg capsules. Three subclasses are recognized, the Digenea, Monogenea, and Aspidogastrea (Fig. 1). The life histories of the Digenea are complex, while those of the Monogenea and Aspidogastrea are simple

Taxonomy. Systematic studies are based mainly on comparative morphology, although life history data are also used In 1957, G. R. LaRue presented an interesting, detailed history of trematode systematics

Though some workers, predominantly the Russians, separate the trematodes into three classes, Trematoda, Udonelloidea, and Monogenoidea, most employ the single class Trematoda, with Digenea and Monogenea as subclasses. It seems certain that these two groups possess ontogenetic and morphological characters which justify subclass rank. The position of the Aspidogastrea is less clear because of their general anatomical similarity to the Digenea and their monogeneid life histories. However, aspidogastreids seem sufficiently different from both groups to deserve equal status (Fig. 2).

The Digenea and Aspidogastrea are probably phylogenetically older than the Monogenea, since they are entoparasitic, parasitize both invertebrates and vertebrates, and have more complicated life histories Since early mollusks were marine forms, these trematodes are probably marine in origin. The Monogenea probably originated in the sea also because a greater divergence of forms has taken place among marine species. All trematodes are strikingly similar to rhadbocoel turbellarians, from which they probably evolved Indeed, some turbellarians have simple sucking holdfasts and tend toward parasitic habits (*see* TURBELLARIA). The major differences between free-living turbellarians and trematodes are the addition of complex holdfasts, development of a cuticle, loss of photoreceptors, and an increase in reproductive capacities with concomitant alterations in the reproductive system, all of which are modifications for the parasitic habit

Morphology. Although the three subclasses share a common internal morphological plan, they differ considerably externally. The Digenea are commonly more elongate, with continuous outlines, while the Monogenea have enlarged anterior and posterior holdfasts and more irregular outlines, and the Aspidogastrea are modified ventrally Digeneids usually have oral and ventral suckers as adult holdfasts Aspidogastreids have large, ventral, adhesive disks with many depressions, or alveoli Monogenea have paired suckers, or adhesive glands, anteriorly and armed suckers or wedges, called opisthaptors, posteriorly Though all subclasses have longitudinal, transverse, and oblique

muscle layers, monogeneids usually have more complex arrangements of longitudinal muscle bands to operate the holdfasts. Digestive systems consist of an anterior mouth, muscular pharynx, short esophagus, and forked blind gut, with the exceptions of *Udonella* and the Aspidogastrea which have single, median digestive sacs. In some digeneids and many monogeneids, the main intestinal branches are much ramified. The nervous system has an enlarged anterior esophageal ganglion or brain. Dorsal nerve trunks usually lacking in Monogenea, as well as ventral, longitudinal, and transverse trunks also occur. Photosensitive eyespots occur in the larvae of many Digenea and Monogenea, and in some adult monogeneids. Digenea usually have a single nephridiopore, while monogeneids possess two Aspidogastrea feature both conditions.

Most trematodes possess turbellarianlike hermaphroditic reproductive systems with single or multiple testes, a single ovary, common gonopore, and a protrusible, or eversible, copulatory organ. A few digeneids, the blood flukes, have separate sexes All three groups may have homologous accessory ducts running from the female tracts to the outside, known as Laurer's canal in the Digenea and Aspidogastrea, or have accessory tracts to the gut, like the genitointestinal canal of the Monogenea Cross-fertilization and self-fertilization are possible and, depending on the availability of mates, both occur, but the former probably predominates. Oviparity is the rule, but a few monogeneids, such as *Gyrodactylus* and *Isancistrinae*, bear living young. The ciliated embryo, which usually does not occur in the Aspidogastridae, is enclosed in an

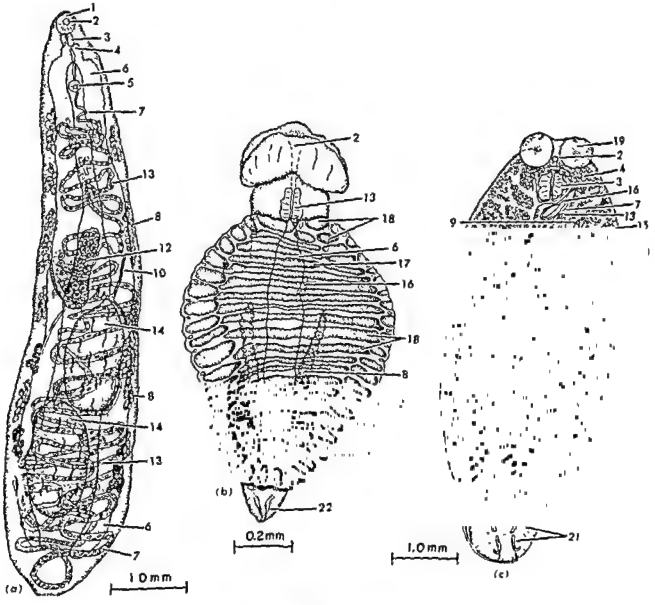


Fig 1. Ventral views of representative trematodes (a) *Haematoloechus*, a digenetic trematode from the lung of a frog. (b) *Cotylaspis ringins*, an aspidogastrid from *Calamus calamus*, the saucer-eye porgy. (c) *Benedenia girellae*, a monogeneid from *Girella nigricans*, the opal-eye. 1, oral sucker; 2, mouth, 3, pharynx; 4, gonopore, 5, ventral sucker or acetabulum, 6, intestine; 7, uterus; 8, vitelline glands; 9, oviduct, 10, vitelloduct; 11, vitelline reservoir, 12, ovary; 13, egg, 14, testis; 15, vas deferens; 16, cirrus; 17, ventral disk or haptor; 18, alveoli; 19, anterior suckers, 20, posthaptor; 21, anchors; 22, nephridiopore

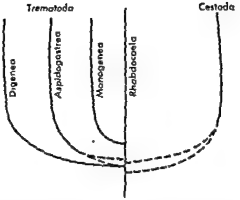


Fig 2 Diagram of the probable lines of evolution of Digenea, Aspidogastrea, and Monogenea from rhabdocoele turbellarians.

operculated egg capsule containing nourishing vitelline cells.

Life history. Egg production of most digeneids is high and, coupled with asexual multiplication of the larval stages, results in enormous numbers of young Most probably succumb to the hazards of their complicated life cycles, therefore, large numbers are necessary The Monogenea produce fewer embryos (Fig 3)

The complicated life cycles of the Digenea involve asexual reproduction in the f...

...develop directly, and find lodgment in the final molluscan host. Some aspidogastreids, however, have an intermediate molluscan, or decapod, host with turtles or fishes as the final host. Monogeneids possess ciliated larvae, which undergo simple metamorphosis into the juvenile form on a single host.

Physiology. Comparatively little is known of helminth physiology. Such studies are complicated by the parasitic habit, small size, and complicated life histories of the worms. Helminth metabolism is not as simple as once believed, but involves great adaptability to varying oxygen tensions and pH ranges, among other factors. Many digeneids are capable of respiration under low oxygen or anaerobic conditions. The physiological activities of larvae and juveniles are probably very different than those of the adults. Trematodes feed upon host tissues, body fluids, and exudates, and both extracellular and intracellular digestion takes place. Transcutaneous absorption of nutrients may occur. Efforts at culturing trematodes on artificial media have met with some success, and several investigators are pursuing transplantation studies

Ecology. Trematodes parasitize a wide variety of invertebrate and vertebrate hosts and occupy almost every available niche within these hosts. The adaptations demanded of the worms for survival are as varied as the characteristics of the microhabitats Over the millions of years of co-evolution of the hosts and their parasites, delicate balances have, for the most part, been attained, and under normal conditions trematodes probably rarely demand more than the host ...
due s
the ar

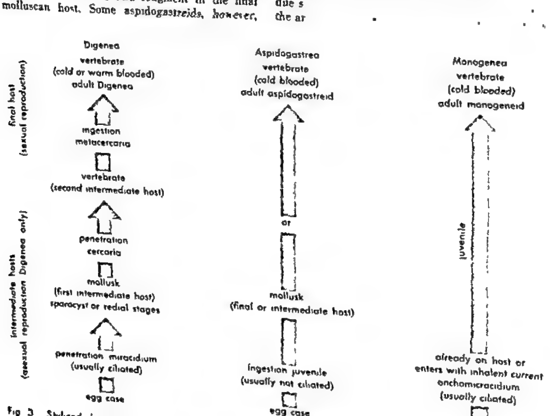


Fig 3 Stylized diagrams of the life histories of the three trematode subclasses

fense reactions of the host. In many entoparasitic trematodes, delicate balances have developed between the protein digestive enzymes of the host and the resistant properties of the cuticle and other tissues of the parasite. In *Paragonimus*, the cuticle actually appears to be digested by the host but is continually renewed by the parasite. Monogeneids do not have the same problems but have had to develop efficient adhesive mechanisms to keep from being swept from the surfaces of their hosts. This struggle has produced many interesting holdfast organs

It seems axiomatic that the host must survive until the parasite can again gain access to another host or until its life cycle is completed. Those parasites which cause the least disruption of the host's activities are probably the oldest as well as the most successful. Immunities are sometimes developed by the hosts. Many trematodes seem to possess such rigid requirements and responses to particular hosts that host-specificity is a phenomenon of considerable significance Monogeneids appear more host-specific than digeneids, and aspidogastreids seem less specific than both. Trematodes are of considerable veterinary and medical importance because, under certain conditions, they cause debility, even death. *See* ASPIDOGASTREA; DIGENEA; MONOGENEA, PLATYHELMINTHES. [W.J HA.]

Bibliography: J G. Baer, *Ecology of Animal Parasites*, 1951; B Dawes, *The Trematoda*, 1946; L H Hyman, *The Invertebrates*, vol. 2, 1951; G. R. La Rue, The classification of digenetic Trematoda: a review and a new system, *Exptl. Parasitol.* 6(3):306–344, 1957.

Trematosauria

A group of Early Triassic labyrinthodonts which, alone of amphibians, appear to have been marine in habitat. The skull is long and pointed; the vertebrae are persistently rhachitomous in structure *See* LABYRINTHODONTIA [A.S R]

Tremolite

The name given to the magnesian-rich monoclinic calcium amphibole $Ca_2Mg_5Si_8O_{22}(OH)_2$. Tremolite forms one end member of a group of solid solution series with iron, sodium, and aluminum The mineral is white to gray, usually exhibiting long prismatic crystals with the 56° amphibole (110) cleavages, and is colorless in thin sections. It is optically negative and has inclined extinction and is thereby distinguished from anthophyllite and cummingtonite, which it resembles Small amounts of iron impart a green color to the mineral; it is then called actinolite, becoming darker green with increasing iron content. The pure mineral is usually the product of metamorphism of impure dolomites and limestones

$$5CaMg(CO_3)_2 + 8SiO_2 + H_2O = Ca_2Mg_5Si_8O_{22}(OH)_2 + 3CaCO_3 + 7CO_2$$

a reaction that occurs in the middle grades of metamorphism. Tremolite in such metamorphosed carbonate rocks is often very pure, free of the usual iron and aluminum content of amphiboles. Minerals ordinarily associated are quartz, calcite, diopside, anthophyllite, talc, and chlorite *See* AMPHIBOLE; ASBESTOS; HORNBLENDE; METAMORPHISM [C.W D]

Trench fever

An infectious, relapsing febrile disease of the Old World caused by bacterialike microorganisms, *Rickettsia quintana*, and transmitted to man by the human body louse, *Pediculus humanus*. H. Mooser and his colleagues in 1954 reported infection in lice in Mexico. Although trench fever was unknown prior to 1915, it involved at least 1,000,000 men during World War I causing, with the exception of influenza, the loss of more man-days than any other disease. It appeared in epidemic form, during World War II, on the eastern European front.

The primary cycle is between man and the louse, and baboons and rhesus monkeys are the only other animals successfully infected. Since organisms have been recovered from the blood as long as eight years after convalescence, and by feeding of clean lice during relapses, man himself probably provides the reservoir.

The agent grows extracellularly in the midgut of the louse but will not grow in tissue cultures or incubating chicken eggs It possesses no antigen in common with *Proteus* strain X (*see* PROTEUS). Complement fixation and agglutination of washed organisms from louse guts or feces by convalescent serums has been reported. Symptomatology varies; accompanied by a sparse macular rash, the first febrile episode may be followed by three to seven fresh bouts over a variable period. Control measures are the same as for epidemic typhus *See* TYPHUS FEVER, EPIDEMIC (LOUSE-BORNE). [C B.P]

Treptostomata

An order of Paleozoic gymnolaematous Bryozoa comprising 10 families and 105 genera which contribute to the formation of many limestone strata (R. Bassler) F. Borg's term, Stenolaemata, for the orders Treptostomata and Cyclostomata is rejected by most bryozoologists Treptostomata lack ovicells and pseudopores and thus differ from the Cyclostomata Treptostomata form lamellate, stemmed, or massive stony colonies, even corallike reefs, and were regarded as corals until 1882 Their apertures are terminal, their zooids long, tubular, calcareous. and partitioned Each tube has a thin, immature section and a thickwalled mature part Special clusters of small zooecia, the monticules and maculae, can be seen on the zoarial surface *See* BRYOZOA [M D.RO]

Trestle

A succession of towers of steel, timber, or reinforced concrete supporting the horizontal stringers of a bridge or other structure It is difficult to distinguish between a trestle and a viaduct, and the terms are used interchangeably by many engineers. A viaduct is defined as a long bri [illegible] a

series of short concrete or masonry spans supported on piers or towers, to carry a road or railroad over a valley, gorge, or another roadway. A viaduct may also be a similar structure of steel girders and towers. It is even more difficult to draw a distinction between a viaduct and a bridge *See* BRIDGE.

The layout for a trestle or a viaduct usually consists of alternate tower spans and free spans For trestles of low height the spans may be supported on bents, each composed of two columns adequately braced in a transverse direction A pair of bents braced longitudinally forms a tower. The columns of one bent of the tower are supported on planed base plates or movable shoes to allow movement *horizontally in the direction* of the longitudinal axis of the trestle Struts connect the column bases and force the movable shoes to slide. The width at the base of a bent is usually not less than one-third the height of the bent. This width is sufficient to prevent *excessive* uplift at windward columns when the trestle is unloaded. *See* STRUCTURES (ENGINEERING). [C.N G.]

Triangle

In the euclidean sense, a triangle is a geometric figure bounded by three (straight) line segments called sides, that meet pairwise in three points called *vertices* It has three angles, each formed by a pair of sides at a vertex, whose three measures have a *sum equal to 180°* or two right angles The sum of two sides of a triangle is greater, and the difference less, than the third side A triangle is called equilateral if *three sides* are equal, isosceles if at least two sides are equal, and *scalene if* no two sides are equal The greater of two sides is opposite the greater angle. A triangle is called acute right, or obtuse, according to whether its largest angle is acute ($<90°$), right ($=90°$), or obtuse ($>90°$)

In a right triangle with sides a, b, c where $a < b < c$, the smallest angle A differs from $172° \cdot a/(b+2c)$ by less than 1′ for $A < 37°$, and by less than 4.5′ for $A < 45°$. Exact relations between the sides and angles of a triangle are the subject of trigonometry. *See* TRIGONOMETRY, PLANE.

Two triangles are said to be congruent if (1) three sides, or (2) two sides and the included angle, or (3) two angles and a side, of the one are equal respectively to the corresponding parts of the other.

A triangle in the projective sense consists of any three nonconcurrent lines and their three points of intersection A spherical triangle is a figure bounded by three great circle arcs on a sphere (*see* SPHERE).

The line segment drawn from a vertex of a triangle to the midpoint of the opposite side is called a median of the triangle. The three *medians* of a triangle intersect in a point—called the *centroid*. *See* GEOMETRY, EUCLIDEAN; POLYGON; PYTHAGOREAN THEOREM. [J.S F.]

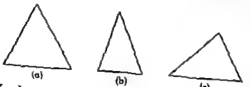


Fig 1 Triangles (a) Equilateral (b) Isosceles (c) Scalene

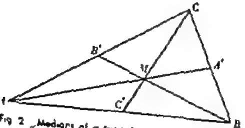


Fig 2 Medians of a triangle AA', BB', and CC'

Triassic

A geological term designating the lowest rock system of the Mesozoic era, proposed by F. von Alberti in 1834 for a sequence of strata in central Germany lying above marine Permian (Zechstein) and below marine Jurassic (Liassic) The name is

PRE-CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	Mississippian	Pennsylvanian	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

derived from the threefold facies division of these strata into a lower nonmarine redbed facies (Bunter), a middle marine limestone, sandstone, and shale facies (Muschelkalk), and an upper nonmarine continental facies similar to the lower division (Keuper). This sequence and character of facies typifies Triassic strata of northern Europe, France, Spain, and northern *Africa*; *it* is commonly known as the Germanic facies. In England the marine Muschelkalk facies is absent In contrast, the predo...

...is commonly called the Alpine facies (Fig 1), and the *sequence of* ammonoid zones in these formations, with the addition from southern Asia of some zones in the lower beds, forms the primary standard sequence of stages in the Triassic system The standard *divisions* and correlation of *the* Germanic and Alpine *units*, plus the main ammonite faunal zones, are shown in the accompanying table.

Standard divisions and correlation of Germanic and Alpine Triassic units

Germanic facies			Alpine facies: Series	Alpine facies: Stage	Alpine facies: Ammonite faunal zones
Keuper	Rhät	Rhétien	Upper Triassic	Rhaetian	*Choristoceras marshi*
Keuper	Gypskeuper	Keuper of French geologists (Marnes irisées)	Upper Triassic	Norian	*Sirenites argonautae* *Pinacoceras metternichi* *Cyrtopleurites bicrenatus* *Cladiscites ruber* *Sagenites giebeli* *Discophyllites patens*
Keuper	Gypskeuper Letten Kohle	Keuper of French geologists (Marnes irisées)	Upper Triassic	Karnian	*Tropites subbullatus* *Carnites floridus* *Trachyceras aonoides* *Trachyceras aon*
Muschelkalk (Calcaire conchylien)			Middle Triassic	Ladinian	*Protrachyceras archelaus* *Protrachyceras reitzi*
Muschelkalk (Calcaire conchylien)			Middle Triassic	Anisian (Virglorian)	*Paraceratites trinodosus* *Paraceratites binodosus* *Nicomedites osmani* *Neopopanoceras haugi*
Bunter or Bundsandstein (Grès bigarré)			Lower Triassic	Scythian (Werfenian)	*Prohungarites similis* *Columbites parisianus* *Tirolites cassianus* *Anasibirites multiformis* *Meekoceras gracilitatus* *Flemingites flemingianus* *Koninckites volutus* *Xenodiscoides fallax* *Prionolobus rosenkrantzi* *Vishnuites decipiens* *Ophiceras commune* *Otoceras woodwardi*

It should be noted from this chart that the French geologists define the term Keuper in a more restricted fashion than do the German geologists, that is, to the middle part of the German sequence (Gypskeuper), and refer the Rhaetian (Rhät) to the Jurassic and the Lettenkohle to the Middle Triassic (Muschelkalk). In the French literature the German terms have been translated, in ascending order, as *Grès bigarré* (Bunter), *Calcaire conchylien* (Muschelkalk), and *Marnes irisées* (Keuper, restricted). Additional but minor variations are the use of Virglorian for Anisian, and Werfenian for Scythian.

Triassic paleogeography. The widespread orogenies which began during the Carboniferous period and continued intermittently until the end of the Paleozoic era brought to an end the pattern of widespread geosynclines that had characterized the Paleozoic. Thus during the later phases of the Paleozoic much of Europe north of the Alps, the Ural mountain region, eastern North America, and much of Asia north of the Himalayas, which had been active geosynclinal regions, were transformed into rigid continental blocks that became part of the stable portion of the respective continents. With the close of the Permian, epicontinental and geosynclinal seas had retreated from most of the continents and the geography of the lands and seas must have been very similar to what it is today. See GEOSYNCLINE, OROGENY.

It is upon this pattern that Triassic history began. Triassic seas were marginal to the continents except for the Tethyan geosyncline extending from the Alps through the Middle East and the Himalayas to Indonesia. Thus we find marine Triassic rocks in all the countries on the margins of the Pacific and marginal to the Arctic Ocean. In many of these areas the marine strata grade landward into continental facies. South America, except for its western margin now occupied by the Andes, was land area with local basins of continental deposition. These deposits, consisting mainly of sandstone, are found in southern Brazil, northern and western Argentina, Uruguay, and Paraguay. Africa was emergent throughout the Triassic except for a narrow coastal belt from Libya through Egypt. In South Africa, great thicknesses of continental sandstones and shales accumulated which are well known for their fossil reptile faunas. Madagascar has mainly continental Triassic formations like those in South Africa except for a few thin marine tongues in the north. Peninsular India was likewise the site of extensive continental deposition which comprises part of the Gondwana series. These deposits consist mainly of sandstones and shales that accumulated in linear, local, grabenlike basins.

Fig 1 Distribution of Triassic outcrops in west-central Europe (From R. C Moore, *Introduction to Historical Geology*, 2d ed , McGraw-Hill, 1958)

Australia was also an *emergent* continental area

[illegible] basins

It was not until the Jurassic that shallow *shelf* seas began to spread beyond the peripheral geosynclinal belts and inundated parts of Eurasia and North America

Triassic life. The great restrictions of epicontinental and geosynclinal seas in the late Paleozoic was accompanied by the extinction of a large part of the marine faunas characteristic of that era *Major taxa* of the corals echinoderms arthropods, mollusks, bryozoans and brachiopods became extinct This great crisis in the evolutionary history of marine animals is reflected in the composition of Triassic marine faunas In the Lower Triassic, for instance, no corals, foraminifera, or bryozoa have been discovered as yet Brachiopods, crinoids, sponges, asteroids, ophiuroids, echinoids, and ostracods are represented by few genera The dominant marine animals in the Early Triassic seas are mollusca, especially the ammonites Pelecypods, nautiloids and gastropods are fairly common but not nearly to the extent of the ammonoids. The most characteristic feature of Early Triassic faunas is their great homogeneity wherever they occur

Chronology and correlation of Triassic marine rocks is based mainly on the sequence and distribution of ammonites There are roughly 400 genera of Triassic ammonoids many of world-wide distribution, in the *Lower Triassic* alone there are 130 genera This reflects a phenomenal evolutionary radiation of the group from a single surviving Paleozoic stock The ammonoids underwent another period of crisis at the end of the Triassic and only a single stock survived into the Jurassic This again reflects another phase of extensive regression of the seas during the Rhaetian.

The great crisis of mass extinction that affected marine invertebrate life at the end of the Paleozoic and led to such a different character of Triassic marine faunas did not affect the animals living on land. There are no sharp breaks or interruptions in the evolutionary history of the amphibians or reptiles, and none in the evolutionary development of plants

In the early part of the Triassic both terrestrial and aquatic vertebrates were comparable to those of the late Paleozoic During the period, however, there were marked changes among both fishes and reptiles and by the end of the Triassic almost every one of the striking vertebrate groups which were to dominate the Jurassic and Cretaceous had appeared

The fossil record of early Mesozoic floras is very poor, but the data available indicate that the flora was greatly impoverished, probably because of unfavorable climates, and consisted mainly of survivors from the Paleozoic. It is not until late in the Triassic that the land plants begin to reflect a distinctly Mesozoic character. From the Late Triassic through the Early Cretaceous, land floras were surprisingly uniform throughout the world This was the period of the great evolutionary radiation of the gymnosperms *See* PALEOBOTANY; PALEOECOLOGY PALEONTOLOGY

Triassic of North America. Geosynclinal and *shelf seas* were confined to the western part of the continent covering western Mexico, the Rocky Mountain and Pacific Coast states, western Canada, and Alaska The Arctic Islands of Canada and northern and eastern Greenland were also periodically inundated by Triassic seas Continental deposits of this age are known in the eastern United States from Florida (subsurface data) to Nova Scotia and in the western United States in Wyoming, Utah Colorado, New Mexico, and Texas.

In the geosynclinal area of western North America a highly complex and varied sequence of Triassic facies many richly fossiliferous is present The western portion of this vast geosynclinal tract is characterized by thick volcanic accumulations, along with a varied suite of sedimentary rocks. The eastern half of this geosynclinal region has no volcanic materials and the sedimentary facies consist of shallow-water limestones, sandstones, and shales that grade eastward into nonmarine redbed formations (Fig 2) This general pattern of facies has persisted since the early Paleozoic At the end of the Early Triassic the seas retreated from the eastern part of the geosynclinal region in the United States and through Middle and Late Triassic time the seas were confined to a region west of a north-south line through central Nevada The area east of this line, which formerly had been covered by geosynclinal seas, became the site of active continental deposition during the Late Triassic and following Jurassic. To illustrate the lithologic character of the strata in the western volcanic part of

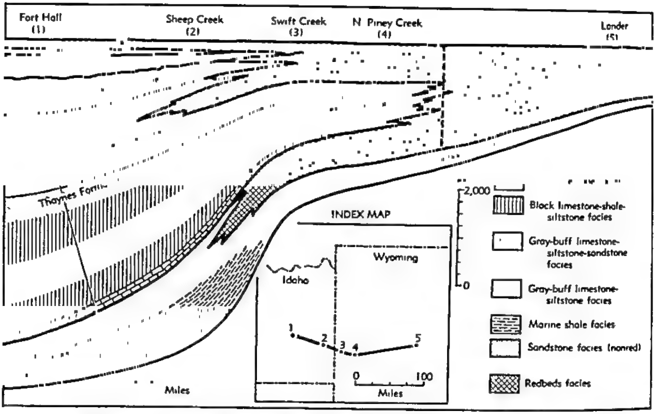


Fig. 2. Triassic deposits of Wyoming and Idaho showing thick marine deposits in the west and continental redbed deposits in the east of equivalent age. (*Data from B. Kummel as used in R C Moore, Introduction to Historical Geology, 2d ed, McGraw-Hill, 1958*)

the geosyncline, the region of southwestern Nevada is classic. The Early Triassic is represented by 3000 ft of marine shales and limestones with tuffaceous sandstones (Candelaria formation) unconformably overlying Permian and older rocks The Middle Triassic is represented by a marine limestone and shale formation (Grantsville) and a volcanic sequence with some lenses of shales and limestones (Excelsior formation) This volcanic formation may be in part equivalent to the Grantsville formation. Resting unconformably upon these formations, the Upper Triassic consists of a varied sequence of marine shales and limestones nearly 2 miles thick which grade without change into the overlying Lower Jurassic.

From California north through southern Alaska no Lower or Middle Triassic is present, only Upper Triassic strata consisting of sedimentary and volcanic rocks. From northern Alaska eastward through the Arctic Islands of Canada to Pearyland there are many areas of Triassic rocks, as yet poorly known except that in one place or another a nearly complete sequence of marine faunas has been observed.

Eastern Greenland has various areas with marine Lower Triassic rocks of a near-shore facies that have yielded a very abundant ammonite fauna, nearly identical with those from the classic Himalayan sections.

Continental Triassic deposits are exposed along the Atlantic Coast and in the Rocky Mountain–Colorado Plateau region. The exposures in eastern North America are confined to a series of isolated troughs extending from Nova Scotia south to North Carolina This is the area of the former Appalachian geosyncline which underwent its final orogenesis late in the Paleozoic. After this orogenic phase a series of downwarped and faulted troughs developed within this new mountain system. Sediments from the adjoining highlands poured into the troughs, forming very thick deposits of sandstones, shales, and conglomerates representing fan, stream, and lake deposits Associated with the sedimentary deposits are igneous rocks in the form of flows, sills, and dikes. These Triassic deposits are known as the Newark Group and are all Upper Triassic in age (Fig. 3). No marine fossils are known from the Newark Group, but land plants, fresh-water fish, and dinosaur tracks are fairly abundant.

A sequence of terrestrial or marginal clastic rocks tentatively classified as Triassic were penetrated in 12 scattered oil-test wells in southeastern Alabama, southwestern Georgia, and north-central Florida. Lithologically, the subsurface sedimentary rocks are closely similar to rocks of the Triassic Newark Group, which crop out at various localities along the Atlantic seaboard from Massachusetts to North Carolina, also, like the Newark Group, intrusions and flows of diabase and basalt cut the Triassic (?) strata in several wells

In the Colorado Plateau area Lower Triassic deposits are termed the Moenkopi formation, to the west and north these grade into marine formations. The Moenkopi consists of dark-red sandstone, siltstone, interbedded gypsum and some marine lime-

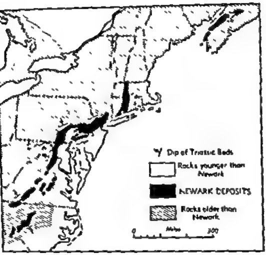


Fig 3 Distribution of Triassic outcrops in eastern North America The rocks occupy structurally depressed belts distributed from North Carolina to Nova Scotia (From R C Moore, *Introduction to Historical Geology*, 2d ed, McGraw-Hill, 1958)

stones in its more western exposures The vertebrate fauna contained in these beds resembles that of the Early Triassic of Spitzbergen and the Bunter of Germany Unconformably overlying the Moenkopi and overlapping it extensively to the east is the Chinle formation, with a basal sandstone or conglomerate often separately recognized as the Shinarump In eastern New Mexico and western Texas, equivalents of the Chinle are termed the Dockum Group From the Wind River Mountains of Wyoming east into South Dakota, extensive Triassic redbeds are designated as the Chugwater and Spearfish formations All of these formations are Upper Tri... faunas of r... formations ... Moenave formation, and possibly the Kayenta formation) are widespread in the southwestern United States

[B KL]

Bibliography: B Kummel, Paleoecology of Lower Triassic formations of southeastern Idaho and adjacent areas, in H S. Ladd (ed.), *Treatise on Marine Ecology and Paleoecology*, Geol Soc Am Mem. 67 vol 2, 1957; B Kummel, in W. J Arkell et al Mollusca, Cephalopoda, Ammonoidea (Triassic ammonoids), *Treatise on Invertebrate Paleontology*, pt L 1957, F H. McLearn, Correlation of the Triassic formations of Canada, *Bull Geol Soc Am*, 64:1205–1228, 1953, J B Reeside, Jr, et al *Correlation of the Triassic formations of North America exclusive of Canada*, *Bull Geol. Soc Am*., 68 1451–1514 1957

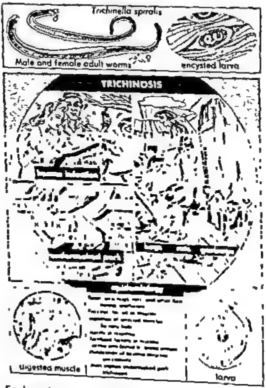
Trichinosis

The presence of *Trichinella spiralis*, a nematode, in the body of man (*see* TRICHUROIDEA) Trichinosis is a serious public health problem and, at times a medical one in North America and Europe The parasite also occurs in many other parts of the world. An estimated 16% of the population of the United States is infected with this parasite

eating encysted larvae in raw or partially cooked pork. Ingested larvae develop in a week into adult nematodes 1.5–4 mm long in the duodenal mucosa. During the 3–16 week period that the adult female parasitizes man, each one discharges 1000–1500 larvae These enter the lymphatics and are carried to all parts of the body where they encyst preferably in active muscles The host reacts by forming a capsule around the larva

Infection by a few larvae may not occasion a noticeable damage. Heavy exposure results in a syndrome of varied nature. The first stage is related to the intestinal phase of parasitism and simulates acute food poisoning, typhoid, or cholera. Skin eruption and respiratory involvement may complicate the picture The second stage coincides with larval migration, that is, inflammatory processes in muscle and circulatory system, swelling or edema, remittent high fever, and hypereosinophilia, which is an abnormally high number of eosinophiles in the blood The third stage relates to encystation of larvae: cachexia, dehydration, blood pressure and neuropathic changes. *See* CHOLERA VIBRIO; FOOD POISONING, BACTERIAL; TYPHOID FEVER

Serious cases are apt to occur from eating smoked but not cooked country sausages Thus the patient's history helps in diagnosis. After two



Epidemiology of trichinosis

weeks serological tests may be of value. *See* SEROLOGY.

Treatment is mainly supportive, that is, it alleviates secondary symptoms such as anemia and malnutrition. Prophylaxis is best accomplished by cooking all pork products. A temperature of 66°C (150°F) kills *Trichinella* larva. Well-cooked pork has a gray or white color.

Curing or pickling pork will not necessarily kill encysted larvae. The larvae will be killed if the pork is kept frozen at −15°C (+5°F) for 21 days or at −30°C (−22°F) for 25 hours *See* PARASITOLOGY, MEDICAL. [J.F.MA.]

Trichloroacetic acid

Trichloroacetic acid, CCl_3COOH, is a colorless, crystalline, deliquescent, highly corrosive acid, with melting point 58°C and boiling point 196–7°C. It is fairly soluble in water and very soluble in alcohol and ether. Trichloroacetic acid is prepared either by nitric acid oxidation of chloral, or by direct chlorination of acetic acid using iodine or phosphorus trichloride as catalyst; it is one of the strongest organic acids known ($K_a = 1.3 \times 10^{-1}$).

Trichloroacetic acid is used as a decalcifier and fixative in microscopy, as a denaturant and precipitant of proteins, and, in the form of dilute solution, in medicine as an astringent and antiseptic. The substance is relatively unstable and decomposes on heating to give chloroform and carbon dioxide When heated in alkaline solutions, it gives carbonates and chloroform. *See* ACETIC ACID; CARBOXYLIC ACID [F B.R.]

Trichomonadida

An order of the class Zoomastigophorea. This group of flagellate protozoans contains several families of uni- or multinucleate species formerly included in the order Polymastigida. Binucleate forms are excluded. Four families comprise this order, the most important being the Trichomonadidae. Several species of this family are found as parasites in vertebrates, including man. H. Kirby erected this order in 1947. *See* POLYMASTIGIDA. [C B C]

Trichomoniasis

Infection with a species of *Trichomonas*, a genus of flagellate protozoa belonging to the class Flagellata. Three species of *Trichomonas* frequently occur in man: *T. tenax* (= *buccalis*) in the mouth, *T. hominis* in the intestine and *T. vaginalis* in the genitourinary tract of both sexes. It is not proved that the first two cause disease.

Vaginitis and urethritis are associated with *T. vaginalis*. The cause of the vaginitis is not single and some unidentified factor or factors are also needed to cause disease. In the United States infection rates in females aged 20–40 vary from about 6% in middle-class groups to 60% in certain closed populations, such as inmates of prisons Defined variables appear to be age (the infection is rare before puberty), multiplicity of sexual partners, and opportunity for treatment.

The infection is often asymptomatic in women Trichomonas vaginitis is characterized by superficial ulcerations and discharge. There is no deep tissue damage, no threat to life, no interference with pregnancy, but the discomfort and painful coitus which may result can be extremely distressing. In males, the infection rarely provokes complaints, the principal sign being a slight urethral discharge. Conjugal infections are not uncommon. Treatment is often, but not invariably, satisfactory. The ideal regimen would avoid reinfection and would eradicate *T. vaginalis* from all its foci in the genitalia, accessory glands, and urinary tract. This is difficult to achieve at the present time *See* MASTIGOPHORA; PARASITOLOGY, MEDICAL. [D W.]

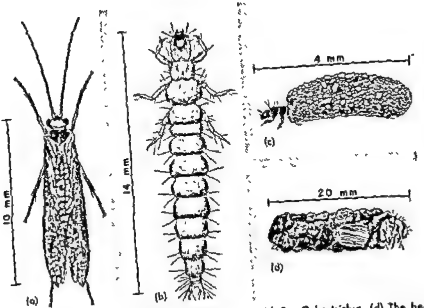
Trichoptera

An aquatic order of the class Insecta commonly known as the caddisflies. The adults have two pairs of well-veined, hairy wings, long antennae, and mouthparts capable of lapping only liquids. The larvae are wormlike, with distinct head, three pairs of legs on the thorax, and a pair of hook-bearing legs at the end of the body. The pupae are delicate, with free appendages held close to the body, and have a pair of sharp mandibles, or jaws, which are used to cut an exit from the cocoon.

The adults live several weeks to several months, the females soon becoming mature. They crawl into water and lay eggs under stones and other objects. The eggs hatch into aquatic larvae. Some larvae construct a fixed retreat and some sort of nest in cracks or crevices Others build a portable case in which to live, and a few build neither, crawling about in moss or under stones Most caddisfly larvae are omnivorous, feeding on algae, other microorganisms, or other aquatic animals which are small enough for them to devour. A few forms are entirely predaceous and live on other aquatic insect larvae When full grown, the free-living, and retreat-making larvae spin an oval cocoon under a rock or in a crevice and pupate in it. The case-makers anchor the case securely to some object in the water and pupate inside it When mature, the pupa, using its jaws, cuts its way out of the cocoon, swims to the surface, sometimes climbing out of the water on a stem or stone, and there the adult emerges from the pupal skin

Except for a brackish-water species in New Zealand and a few moss-inhabiting species in Europe, caddisflies occur only in fresh water They abound in cold or running water relatively free from pollution. Altogether they comprise a large and important segment of the biota of such habitats and of the fish food economy

The Trichoptera include about 7000 described species, comprising 26 families, and occur in practically all parts of the world The order probably arose over 200,000,000 years ago in early Mesozoic time, for fossils of typical trichopteran wings have been found in late Triassic deposits Many existing genera are probably of Cretaceous origin, hence it seems almost certain that representatives of all the



(a) Adult and (b) the free-living larva of the widespread genus *Rhyacophila*. (c) The head and thorax of the larva protruding from the purselike case of a ... *Ochrotrichia*. (d) The head and thorax ... e caddis- ... y)

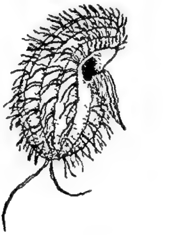
diverse family lines evolved during the middle part of the Mesozoic era at the same time that the dinosaurs were proliferating The Trichoptera were originally cool-adapted animals, as most of the primitive forms still are, but many warm-adapted lines have evolved As a result, the caddisflies are found in arctic, temperate, and tropical habitats *See* INSECTA [H H.R]

Bibliography. C. Betten, *The Caddis Flies or Trichoptera of New York State*, N.Y State Mus. ... B 11 ... 1921 H H Ross *The Caddisflies*, ...

1956, G Ulmer, Trichoptera, *Genera Insectorum*, 60 1–259, 1907, R. L. Usinger (ed) *Aquatic Insects of California*, 1956

Trichostomatida

An order of the Holotricha comprising a small group of ciliates whose species show some advance in structural complexity over the gymnostomes. No true buccal ciliature is present, but a vestibulum, a passageway leading from the outside of the body into the cytostome or true cell-mouth, is characteristically present These animals exist in a wide range of habitats A few forms live in cattle, sheep, and horses, but cause no harm to these hosts *Colpoda* (see illustration) is a well known genus that has been studied widely by experimentalists as well as taxonomists *Pseudoprorodon* is a primitive trichostome and *Balantidium*, until recently mistakenly considered to be a member of the quite advanced spirotrichous order Heterotrichida, has received considerable attention This last mentioned genus is of some medical significance as it contains the sole ciliated protozoan species parasitic in man *Balantidium coli* lives in the large intestine and is the causative agent of the fairly uncommon disease balantidiosis. *Paramecium* is almost universally listed as a trichostome, but in reality, belongs among the hymenostomes. *See* HOLOTRICHA, *see also* BALANTIDIASIS [J O.C.]



Colpoda, a trichostomatid; length 15–40 microns

Trichroism

When certain optically anisotropic transparent crystals are subjected to white light, a cube of the material is found to transmit a different color through each of the three pairs of parallel faces. Such crystals are sometimes termed trichroic, and the phenomenon is called trichroism. This expression is used only rarely today since the colors in a particular crystal can appear quite different if the cube is cut with a different orientation with re-

spect to the crystal axes. Accordingly, the term is frequently replaced by the more general term pleochroism. Even this term is being replaced by the phrase linear dichroism or circular dichroism to correspond with linear birefringence or circular birefringence (*see* BIREFRINGENCE; DICHROISM).

Cordierite is a typical trichroic crystal. In light with a vibration direction parallel to the X axis of the index ellipsoid the crystal appears yellow. With the vibration direction parallel to the Y axis the crystal is dark violet. In the Z direction the crystal is clear.

The phenomena of trichroism can be explained crudely as follows. Classically one can consider an electron in a biaxial crystal as having three different force constants associated with a displacement directed along each of the principal axes. Linear polarized light traveling along the X axis with its electric vector parallel to the Y axis will displace the electron against the Y force constant and will experience a certain absorption and retardation. It will be unaffected by the force constants in the X and Z directions. Similarly, polarized light traveling in the Y direction will experience absorption and retardation. Unpolarized light will also be absorbed in a different fashion depending on the direction of propagation In this case light traveling in the X direction can be considered as composed of an equal mixture of light polarized parallel to the Y axis and the Z axis. The absorption will be intermediate between the two polarization directions. *See* CRYSTAL OPTICS; POLARIZED LIGHT [B H BI]

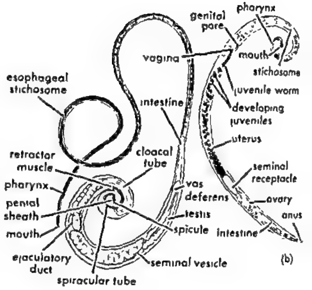
Bibliography: E. S. Larsen and H Berman, *The Microscopic Determination of the Nonopaque Minerals*, USGS Bull. 848, 1934; F Pockels, *Lehrbuch der Kristalloptik*, 1904.

Trichuroidea

A group of nematodes parasitic in various vertebrates and characterized by the peculiar structure of the portion of the body which contains the esophagus. This group is given the status of an order or superfamily in the class Nematoda, according to the views of the specialist The life cycle varies greatly in different species. In some it is direct, in others a transport or an intermediate host may be involved The group includes disease-producing parasites of man, domestic mammals, and poultry.

Common examples. Two genera are common parasites of man; *Trichuris* and *Trichinella* Others commonly occur in domestic animals.

Trichuris trichiura. This roundworm (Fig. 1*a*), a common parasite of man in warm climates, belongs to a group known as whipworms because of the short thick posterior portion of the body and the long slender anterior part The adult worm embeds in the anterior portion in the intestinal wall of the host. The eggs, which have a heavy shell of characteristic appearance, are passed in the feces In warm, moist soil larvae develop within the eggs to the infective stage in a few weeks, but the eggs



(a)

Fig 1. (a) Trichuris trichiura, adult male, 30–50 mm (b) Trichinella spiralis, adult female (From F. A. Brown, Selected Invertebrate Types, Wiley, 1950)

hatch only after ingestion by the host. Human infections result from hand to mouth transfer of eggs or from eating contaminated vegetables

Trichinella spiralis This nematode (Fig 1*b*) is an important cause of human disease and is a common parasite of pigs. In either host the adult worms live in the intestine where the females burrow into the wall and deposit larvae into the blood stream in large numbers over a period of weeks Carried to all parts of the body, the larvae enter the muscle fibers causing fever and muscular pain until they become encysted (Fig 2) In pigs, which become infected from eating pork scraps in garbage, and in humans who are infected from eating undercooked pork, the larvae are digested from the cysts and develop into adult worms Severity of symptoms is proportional to the number of larvae eaten, relatively few cases result in death

Other species. Other worms in the group are of veterinary importance while still others are of great biological interest Some species of *Capil*

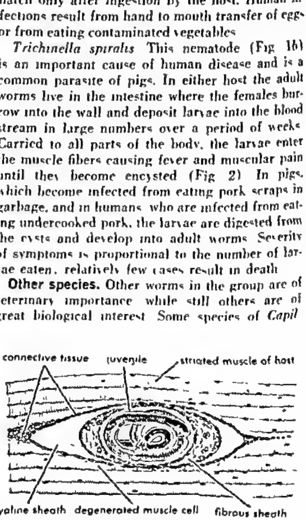


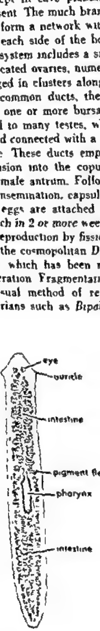
Fig 2 Trichinella spiralis, encysted juvenile, female 3–4 mm, male 1.5 mm. Cyst shown in section, juvenile shown as an intact specimen. (From F A Brown, Selected Invertebrate Types, Wiley, 1950)

laria have a direct cycle while others involve transport or intermediate hosts. *Trichosomoides* is interesting because the diminutive male lives in the vulva or uterus of the larger female. *See* NEMATODA; TRICHINOSIS; WHIPWORM DISEASE. [J.A.S.]

Bibliography: D. L. Belding, *Textbook of Clinical Parasitology*, 2d ed., 1952; S. E. Gould, *Trichinosis*, 1945.

Tricladida

An order of the Turbellaria which are several millimeters to 50 or more centimeters in length. They have diverticulated intestines with a single anterior branch and two posterior branches separated by the plicate pharynx or pharynges. Rhabdites are numerous and except in cave planarians two to many eyes are present. The much branched protonephridial tubules form a network with numerous nephridiopores on each side of the body. The female reproductive system includes a single pair of small anteriorly located ovaries, numerous minute yolk glands arranged in clusters along either side of the body, the common ducts, the female antrum, and usually one or more bursae. The male system has several to many testes, which are lateral in position and connected with a single sperm duct on each side. These ducts empty either directly or after fusion into the copulatory organ which lies in the male antrum. Following copulation and mutual insemination, capsules containing several fertilized eggs are attached to objects in the water and hatch in 2 or more weeks into young worms. Asexual reproduction by fission is common in forms such as the cosmopolitan *Dugesia tigrina* (see illustration) which has been much used in studies on regeneration. Fragmentation and regeneration is the usual method of reproduction in some land planarians such as *Bipalium kewense*, an exotic species which has become established through much of the southern United States. The marine planarian *Bdelloura candida* is a commensal on the horseshoe crab. *See* TURBELLARIA. [E.R.J.]

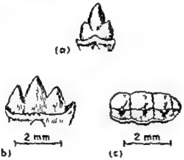


Dugesia tigrina

Bibliography: L. von Graff, *Tricladida*, in H. G. Bronn (ed.), *Klassen und Ordnungen des Tier-Reichs*, vol. 4, pt. 2, 1912–1917.

Triconodonta

The triconodonts were small primitive carnivores or insectivores. They have been found in the Middle and Upper Jurassic of England and the Upper Jurassic and Middle Cretaceous of North America. The molars had three main cusps arranged in a longitudinal series. An internal cingulum (a ledge running along the side of the tooth near the base of the crown) was present on the lower molars. A weak internal and a strong external cingulum were present on the upper molars. The dental formula of 4 incisors, 1 canine, 4 premolars, and 5 molars was probably the primitive condition of the triconodonts. An angular process was not developed on the lower jaw. *See* CARNIVORA FOSSILS; INSECTIVORA FOSSILS.



(a) External view of a lower molar of Amphilestes, an amphilestine triconodont. (b) Internal and (c) occlusal views of a lower molar of Priacodon, a triconodontine triconodont. (After G. G. Simpson, 1929)

Triconodonts were abundant in Middle and Late Jurassic time, but by the Middle Cretaceous they had become quite rare and were probably extinct by the end of the Mesozoic. The medial cusp of the molars of the Middle and Late Jurassic amphilestine triconodonts was larger than the other cusps, but in the Late Jurassic and Middle Cretaceous triconodontine triconodonts the three main cusps were of approximately equal size (see illustration). The lower molars evidently sheared obliquely with the internal side of the upper molars. The lower molars were locked together by a spur on the posterior end of a tooth which fitted into a groove on the anterior end of the tooth behind.

The ancestry of the triconodonts is uncertain, but they may be distantly related to the primitive docodont *Morganucodon*. The known triconodonts probably did not give rise to any other order of mammals. *See* DOCODONTA. [W.A.CL.]

Trifluoroacetate

One of two types of compounds derived from trifluoroacetic acid, $HC_2F_3O_2$ or CF_3COOH. One type is obtained by the reaction of trifluoroacetic acid and bases to give salts that contain the negative trifluoroacetate ion, CF_3COO^-, for example, sodium trifluoroacetate.

The second type is an ester derived from trifluoroacetic acid and an alcohol, for example, ethyl trifluoroacetate, $CF_3COOC_2H_5$.

Sodium trifluoroacetate may be used as an intermediate in the synthesis of other fluorinated compounds. *See* Carboxylic acid; Halogenated hydrocarbon. [E E WR.]

Trigger circuit

An electronic circuit that generates or modifies an existing waveform to produce a pulse of short time duration with a fast-rising leading edge. This waveform, or trigger, is normally used to initiate a change of state of some relaxation device, such as a multivibrator. The most important characteristic of the waveform generated by a trigger circuit is the fast leading edge. The exact shape of the falling portion of the waveform is of secondary importance, although it is important that the total duration time is not too great. A pulse generator such as a blocking oscillator may also be used and identified as a trigger circuit if it generates sufficiently short pulses. *See* Blocking oscillator; Pulse generator

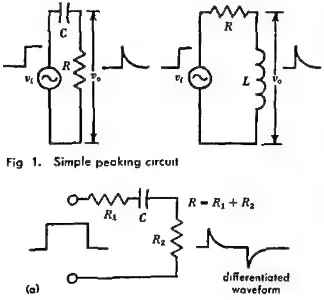


Fig 1. Simple peaking circuit

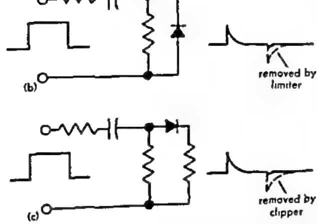


Fig. 2. Differentiated pulses. (a) Basic circuit. (b) Limiting unwanted portion. (c) Clipping unwanted portion.

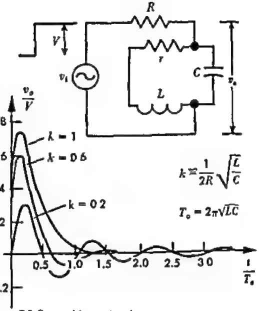


Fig 3. *RLC* peaking circuit.

Peaking (differentiating) circuits. These circuits, which accent the higher-frequency components of a pulse waveform, cause sharp leading and trailing edges and are therefore used as trigger circuits. The simplest form of peaking circuits are the simple *RC* and *RL* networks shown in Fig 1. If a steep wavefront of amplitude V is applied to either of these circuits, the output will be a sudden rise followed by an exponential decay according to the equation

$$v_o = V\epsilon^{-kt} \tag{1}$$

where $k = 1/RC$ or R/L

These circuits are often called differentiating circuits because the outputs are rough approximations of the derivative of the input waveforms, if the *RC* or *R/L* time-constant is sufficiently small.

If a pulse is applied to the differentiating circuit, the resultant waveform shown in Fig 2 may be used as a trigger. It is sometimes necessary, however, to remove by limiting or clipping the undesired portion of the waveform to prevent circuits from responding to it.

The *RL* circuit of Fig 1 cannot be considered in its simplest form when extremely fast rise-times are required because of the distributed capacitance and small series resistance associated with the inductance. A more accurate representation of the circuit is that of Fig. 3. The response is limited as shown for a fixed value of L and C. The value for $k = 1$ is referred to as critical damping. A value of k slightly less than unity provides a pulse that is a suitable trigger for many applications.

Ringing circuits. A circuit of the form shown in Fig. 3 that is highly underdamped, or oscillatory ($k \gg 1$), and is supplied with a step or pulse input is often referred to as a ringing circuit. When used in the output of a vacuum tube or transistor as in Fig 4, this circuit can be used as a trigger circuit.

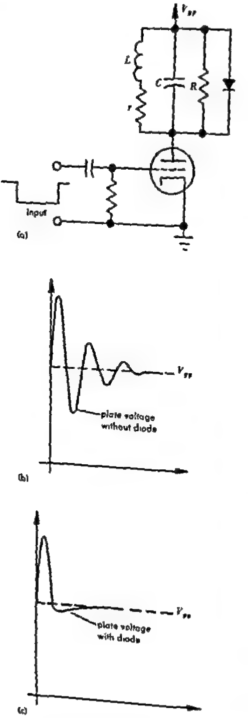


Fig 4 Ringing circuit as trigger source (a) Circuit diagram (b) Plate-voltage waveform without diode limiter (c) Plate-voltage waveform with diode limiter.

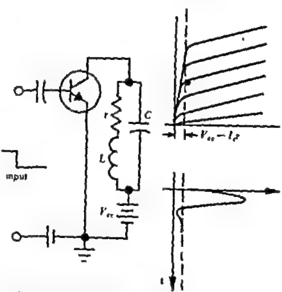


Fig 5 Ringing circuit with transistor saturation damping

When the input pulse is applied, current in the plate circuit is immediately cut off. Since the current in L cannot change instantaneously, it flows in the LC circuit in an oscillatory manner, gradually decaying because of the resistance in the circuit. However, if the diode is in the circuit, the circuit will be highly overdamped for the negative portion of the oscillatory waveform and the oscillations will be damped ...

op ...

wh... transistor goes into saturation as shown in Fig. 5. The diode is therefore not required. For other waveforms *see* WAVE-SHAPING CIRCUITS.

[G.M.G.]

Bibliography. J. Millman and H. Taub, *Pulse and Digital Circuits*, 1956.

Triglyceride

A simple lipid. Triglycerides are fatty acid triesters of the trihydroxy alcohol glycerol which are present in plant and animal tissues, particularly in the food storage depots either as simple esters in which all the fatty acids are the same or as mixed esters in which the fatty acids are different. The triglycerides constitute the main component of natural fats and oils

The generic formula of a triglyceride is

$$\begin{array}{l} CH_2\text{—OOC—R} \\ | \\ CH\text{—OOC—R'} \\ | \\ CH_2\text{—OOC—R''} \end{array}$$

where RCO_2H, $R'CO_2H$ and $R''CO_2H$ represent molecules of either the same or different fatty acids, such as butyric or caproic (short chain), palmitic or stearic (long chain), oleic, linoleic, or linolenic (unsaturated). Saponification with alkali

releases glycerol and the alkali metal salts of the fatty acids (soaps). The triglycerides in the food storage depots represent a concentrated energy source, since oxidation provides more energy than an equivalent weight of protein or carbohydrate.

Animal and vegetable triglycerides contain predominantly even-chain-length fatty acids, with palmitic and oleic acids as the main components. Since n fatty acids may be esterified in $(n^3 + n^2)/2$ ways into glycerol, and since natural fats contain a variety of fatty acids, the number of component triglycerides of a relatively simple natural fat or oil may be high. Some pure simple and mixed triglycerides have been isolated from natural fats by fractional crystallizations at low temperatures, but in general physical methods are not yet available for the separation of naturally occurring mixtures. Several theories such as those of even distribution and partial random distribution have been advanced to account for the distribution of the fatty acids in the triglycerides. Many synthetic triglycerides have been prepared, and the study of the physical properties of these compounds has provided much useful information. Melting-point, x-ray–diffraction, and infrared-spectroscopy investigations have shown that triglycerides may exist *in at least three polymorphic modifications. See* MOLECULAR STRUCTURE AND SPECTRA.

The physical and chemical properties of fats and oils depend on the nature of the fatty acids present. Saturated fatty acids give higher-melting fats and represent the main constituents of solid fats, for example, lard and butter. Unsaturation lowers the melting point of fatty acids and fats. Thus, in the oils of plants, unsaturated fatty acids are present in large amounts, for example, oleic acid in olive oil and linoleic and linolenic acids in linseed oil. Oils are hydrogenated commercially to produce the proper consistency and melting point for use as edible fats. *See* CARBOXYLIC ACID, FAT AND OIL, EDIBLE; LIPID. [H.E.C; R.H.C.]

Bibliography: W. R. Bloor, *Biochemistry of the Fatty Acids*, 1943; T. P. Hilditch, *The Chemical Constitution of Natural Fats*, 2d ed., 1947; K. S. Markley, *Fatty Acids*, 1947; A. W. Ralston, *Fatty Acids and Their Derivatives*, 1948.

Trigonometric curve

The graphical representation of y as a trigonometric function of θ. To obtain the graphs of

$$y = a \sin \theta$$

for the interval $0 \leqq \theta \leqq 2\pi$, prepare the accompanying *table of values of trigonometric functions*, plot the corresponding points, and connect them by a smooth curve. This gives the graph of Fig. 1. Since $\sin(\theta + k2\pi) = \sin \theta$ where k is an integer (*see* TRIGONOMETRY, PLANE), the complete graph consists of an endless repetition both to the right and left of the curve shown.

Since $\sin \theta = \cos(\theta - \frac{1}{2}\pi)$, Fig. 1 is also the graph of $y = a \cos(\theta - \frac{1}{2}\pi)$. The graph of $y = a \cos \theta$ has the same shape as that of $y = a \sin \theta$

Values of trigonometric functions

θ (radians)	$y = a \sin \theta$	θ (radians)	$y = a \sin \theta$
0	0	$\frac{7}{6}\pi$	$-0.5a$
$\frac{1}{6}\pi$	$0.5a$	$\frac{8}{6}\pi$	$-0.87a$
$\frac{2}{6}\pi$	$0.87a$	$\frac{9}{6}\pi$	$-1a$
$\frac{3}{6}\pi$	a	$\frac{10}{6}\pi$	$-0.87a$
$\frac{4}{6}\pi$	$0.87a$	$\frac{11}{6}\pi$	$-0.5a$
$\frac{5}{6}\pi$	$0.5a$	$\frac{12}{6}\pi$	0
$\frac{6}{6}\pi$	0		

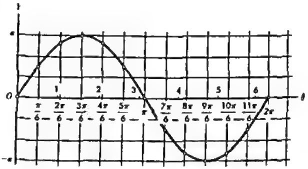


Fig. 1 $y = a \sin \theta$.

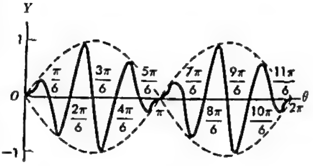


Fig. 2. $y = \sin \theta \sin 6\theta$

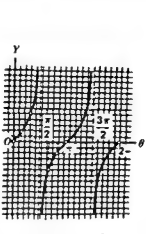


Fig. 3. $y = \tan \theta$.

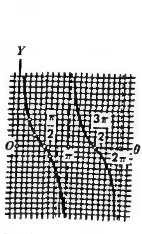


Fig. 4 $y = \cot \theta$

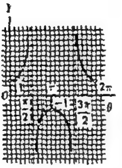


Fig 5 $y = \sec\theta$.

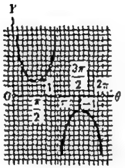


Fig 6. $y = \csc\theta$.

and could be obtained from Fig. 1 by translating the curve $\frac{1}{2}\pi$ units leftward.

If each number on the θ axis were divided by m, Fig 1 would be the graph of $y = a \sin m\theta$.

Since Fig 1 shows one complete set of values of $y = a \sin\theta$ for values of θ from 0 to 2π, the period of $\sin\theta$ is 2π. Also, from the graph of $y = a \sin m\theta$, it can be shown that *the period* of $\sin m\theta$ is $2\pi/m$.

The full line curve of Fig 2 shows the graph of $y = \sin\theta \sin 6\theta$. It indicates roughly the situation of radio waves where a low frequency audio wave is impressed upon a high frequency wave.

Figures 3, 4, 5 and 6 show the graphs of $y = \tan\theta$, $y = \cot\theta$, $y = \sec\theta$, and $y = \csc\theta$, respectively.

[L.M.K.]

Trigonometry, plane

The functions of trigonometry are extremely important because they may be used to represent ranges of values which are repeated again and again. They represent periodic phenomena such as the *motions* of pendulums or the *analysis of* alternating current electricity. In fact they play a basic role in the theories of light, sound, radio, television, and generally in all *phenomena* of a vibratory character.

Also, trigonometry is used to find a vast network of lengths that cannot be measured directly. Surveyors *use it to find* heights of mountains, distances across lakes and countries, and positions of places, engineers use it in the design of large structures and roads, astronomers use it in accurate me[illegible]

[illegible]tions.

Angles. Unless otherwise specified it is understood that all figures considered in this article lie in a plane. If a half-line or *ray*, having end point O (see Fig. 1) rotates about O from an initial position OA to terminal position OB, it is said to generate the *angle AOB*. A rotation of $\frac{1}{360}$ part of a complete rotation about a point is called a degree and written 1°. One-sixtieth of a degree is called a minute and written 1′. Generally angles generated by rotation in a clockwise direction are called negative angles, and those generated by counterclockwise turning are called positive angles. Figure 2, indicating various angles, illustrates this concept.

A radian (see Radian measure) is an angle subtended at the center of a circle by an arc of the circle equal in length to its radius (see Fig. 3). It is used in purely theoretical discussions because it avoids cumbersome constants.

Trigonometric functions. Figure 4 shows an angle θ with vertex O at the origin of a system of rectangular coordinates (*see* Coordinate, systems, graphical) and with initial line OA directed in the positive direction of the x axis. Let P, any point except O, on the terminal ray of the angle θ, have coordinates x and y and be at a distance r from O. Then the six functions, sine, cosine, tangent, cotangent, secant, and cosecant of θ, respectively abbreviated by $\sin\theta$, $\cos\theta$, $\tan\theta$, $\cot\theta$, $\sec\theta$, $\csc\theta$, are defined by

$$\sin\theta = y/r, \cos\theta = x/r, \tan\theta = y/x \quad (1)$$

$$\csc\theta = r/y, \sec\theta = r/x, \cot\theta = x/y \quad (2)$$

where r is positive. These ratios are independent of the position of the point (x,y) on the terminal ray.

Figure 4 shows that any angle with initial side OA obtained by adding an integral multiple of 360° to θ has the same terminal ray as θ, and therefore that

$$\text{fn}\ (\theta + k360°) = \text{fn}\ (\theta) \quad (3)$$

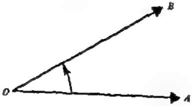


Fig. 1

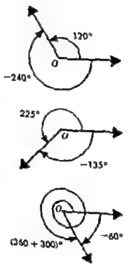


Fig 2

where fn stands for sin, cos, tan, cot, sec, or csc, and k is an integer. In other words the trigonometric functions are periodic, and a period is 360° or 2π radians. This is the property that makes them so useful in theoretical considerations involving periodic values.

Figure 5 shows four angles 60°, 120°, 240°, and 300° having terminal lines containing respective points $(1,\sqrt{3})$, $(-1,\sqrt{3})$, $(-1,-\sqrt{3})$, and $(1,-\sqrt{3})$. Since $r = \sqrt{1^2 + (\sqrt{3})^2} = 2$ in all cases the values from Eq. (1) can be tabulated as follows:

	sin	cos	tan	cot	sec	csc
60°	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}$	$\sqrt{3}$	$\frac{1}{3}\sqrt{3}$	2	$2/\sqrt{3}$
120°	$\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}$	$-\sqrt{3}$	$-\frac{1}{3}\sqrt{3}$	-2	$2/\sqrt{3}$
240°	$-\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}$	$\sqrt{3}$	$\frac{1}{3}\sqrt{3}$	-2	$-2/\sqrt{3}$
300°	$-\frac{1}{2}\sqrt{3}$	$\frac{1}{2}$	$-\sqrt{3}$	$-\frac{1}{3}\sqrt{3}$	2	$-2/\sqrt{3}$

(4)

In all cases $|x| \leqq r$, $|y| \leqq r$, and therefore the range of the sine and of the cosine is -1 to $+1$ inclusive. Figure 5 shows the angles 0°, 90°, 180°, and 270° containing on their terminal rays the respective points (1,0), (0,1), (−1,0) and (0,−1). Using Fig. 5 and Eqs. (1) and (3), the values may be tabulated as follows where k is an integer:

	0 + k360°	90° + k360°	180° + k360°	270° + k360°
sin	0	1	0	−1
cos	1	0	−1	0
tan	0	Undefined	0	Undefined

(5)

If the angle θ is near to, but less than 90°, x is small compared to y and therefore the tangent is large; in fact $\tan\theta$ ranges through all positive numbers as angle θ increases from 0° to 90°. The graphs (*see* TRIGONOMETRIC CURVE) of the trigonometric functions give a clear picture of the ranges and general behavior of these functions. The three functions shown in Eq. (2) are comparatively unimportant. The cotangent is the reciprocal of the tangent, $\cot\theta = 1/\tan\theta$, and it takes on all positive values as θ varies from 0° to 90°. Also,

$$r \geqq |x| \qquad r \geqq |y|$$

therefore

$$|\sec\theta| \geqq 1 \qquad |\csc\theta| \geqq 1$$

From Eqs. (1) and (2), it follows that

$$\csc\theta = \frac{1}{\sin\theta} \qquad \sec\theta = \frac{1}{\cos\theta} \qquad \cot\theta = \frac{1}{\tan\theta} \tag{6}$$

$$\tan\theta = \frac{\sin\theta}{\cos\theta} \tag{7}$$

Also from Fig. 4, $y^2 + x^2 = r^2$. Dividing this through by r^2 and using Eq. (1), $(\sin\theta)^2 + (\cos\theta)^2 = 1$, which is written

$$\sin^2\theta + \cos^2\theta = 1 \tag{8}$$

Using Eqs. (6), (7), (8), and others mentioned below, many complicated expressions can be transformed to simple ones.

Addition formulas. The addition formulas of trigonometry, namely

$$\cos(\phi + \theta) = \cos\phi\cos\theta - \sin\phi\sin\theta \tag{9}$$

$$\sin(\phi + \theta) = \sin\phi\cos\theta + \cos\phi\sin\theta \tag{10}$$

lead to practically all of the relations between the trigonometric functions.

Figure 6 is used in the proof of Eqs. (9) and (10). It represents a set of coordinate axes and a circle O of radius 1 with center at (0,0). Points A (1,0), B, C, D are so placed on circle O that angle AOB is an angle α, and COD is angle AOB revolved through angle θ about (0,0). Hence B is the point $(\cos\alpha, \sin\alpha)$, C is $(\cos\theta, \sin\theta)$, and D is $[\cos(\alpha+\theta), \sin(\alpha+\theta)]$. Using the distance formula (*see* ANALYTIC GEOMETRY) to express the fact that chord CD equals chord AB,

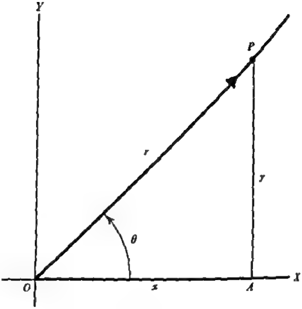


Fig. 4.

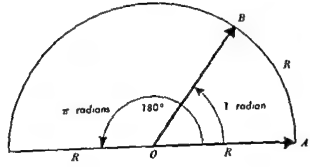


Fig. 3

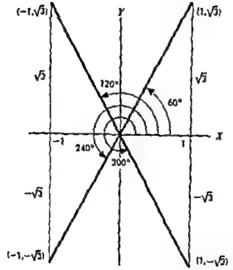


Fig 5.

$$\cos^2(\alpha+\theta) - 2\cos(\alpha+\theta)\cos\theta + \cos^2\theta$$
$$+ \sin^2(\alpha+\theta) - 2\sin(\alpha+\theta)\sin\theta$$
$$+ \sin^2\theta = 1 - 2\cos\alpha + \cos^2\alpha + \sin^2\alpha \quad (11)$$

Substituting Eq. (8) in this three times, once for $\theta = \alpha + \theta$, once for $\theta = \theta$, and once for $\theta = \alpha$, and simplifying slightly,

$$\cos(\alpha+\theta)\cos\theta + \sin(\alpha+\theta)\sin\theta = \cos\alpha \quad (12)$$

Formula (12) holds for any values of angles α and θ In Eq (12), α may be replaced by $(\phi - \theta)$ to obtain

$$\cos(\phi-\theta) = \cos\phi\cos\theta + \sin\phi\sin\theta \quad (13)$$

It is surprising and pleasing to derive with ease a long list of the useful formulas of trigonometry from Eqs (13) and (5). In Eq (13), ϕ may be replaced by 90°, and Eq (5) used to obtain

$$\cos(90° - \theta) = \sin\theta \quad (14)$$

Equation (14) shows that $\cos[90° - (-\theta)] = \sin(-\theta)$ and in combination with Eqs. (13) and (5) shows that $\cos[\theta - (-90°)] = -\sin\theta$; therefore

$$\sin(-\theta) = -\sin\theta \quad (15)$$

In Eq (13), ϕ may be replaced by 0 and, making use of Eq. (3),

$$\cos(-\theta) = \cos\theta \quad (16)$$

In Eq (13) θ may be replaced by $-\theta$, and, using Eqs (15) and (16), Eq (9) is obtained In Eq. (14) θ may be replaced by $(90° - \phi)$ to get

$$\cos\phi = \sin(90° - \phi) \quad (17)$$

In Eq (13) ϕ may be replaced by $(90° - \phi)$, and, using Eqs (14) and (17), Eq (10) is obtained In Eq (10) θ may be replaced by $-\theta$, and using Eqs (15) and (17),

$$\sin(\phi-\theta) = \sin\phi\cos\theta - \cos\phi\sin\theta \quad (18)$$

Reduction formulas. The equations numbered (19) to (24) together with Eq (5) make it possible to express a trigonometric function of any angle as a function of a positive angle no greater than 45°. Tables giving the values of the trigonometric functions at convenient intervals have been computed. Because of these reduction formulas, it is necessary to compute their values only for the interval from 0° to 45°. Using Eqs (9), (10), (13), and (18) together with Eq (5), the following reduction formulas may be verified:

$$\sin(90° \pm \theta) = \cos\theta,\ \cos(90° \pm \theta) = \mp\sin\theta \quad (19)$$
$$\tan(90° \pm \theta) = \mp\cot\theta,\ \cot(90° \pm \theta) = \mp\tan\theta \quad (20)$$
$$\sin(180° \pm \theta) = \mp\sin\theta$$
$$\cos(180° \pm \theta) = -\cos\theta \quad (21)$$
$$\tan(180° \pm \theta) = \pm\tan\theta$$
$$\cot(180° \pm \theta) = \pm\cot\theta \quad (22)$$
$$\sin(270° \pm \theta) = -\cos\theta$$
$$\cos(270° \pm \theta) = \pm\sin\theta \quad (23)$$
$$\tan(270° \pm \theta) = \mp\cot\theta$$
$$\cot(270° \pm \theta) = \mp\tan\theta \quad (24)$$

For example by Eqs. (3) and (24)

$$\tan(1745°) = \tan(1745° - 4 \quad 360°) = \tan 305°$$
$$= \tan(270° + 35°) = -\cot 35°$$

Double angle and half angle formulas. Substituting θ for ϕ in Eqs. (9) and (10),

$$\sin 2\theta = 2\sin\theta\cos\theta \quad (25)$$
$$\cos 2\theta = \cos^2\theta - \sin^2\theta \quad (26)$$

Solving Eq. (26) and $\cos^2\theta + \sin^2\theta = 1$ for

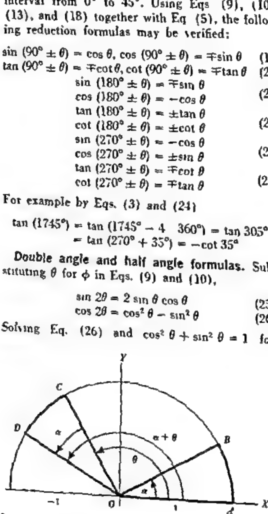


Fig 6

Fig 7

$\sin\theta$ and $\cos\theta$, and replacing θ by $\frac{1}{2}\theta$ in the result,

$$\sin \tfrac{1}{2}\theta = \pm\sqrt{(1-\cos\theta)/2} \tag{27}$$

$$\cos \tfrac{1}{2}\theta = \pm\sqrt{(1+\cos\theta)/2} \tag{28}$$

Dividing Eq. (27) by Eq. (28) member by member,

$$\tan \tfrac{1}{2}\theta = \pm\sqrt{\frac{1-\cos\theta}{1+\cos\theta}} \tag{29}$$

Conversion formulas. The equations numbered (30) to (33) below are used to transform a product into a sum, and Eqs. (35) to (38) are used to transform a sum into a product. This is generally done for some purpose such as evaluation by logarithms or by a computing machine.

First adding and then subtracting Eqs. (9) and (13), member by member,

$$\cos(\phi+\theta) + \cos(\phi-\theta) = 2\cos\phi\cos\theta \tag{30}$$

$$\cos(\phi+\theta) - \cos(\phi-\theta) = -2\sin\phi\sin\theta \tag{31}$$

Similarly from Eqs (10) and (18),

$$\sin(\phi+\theta) + \sin(\phi-\theta) = 2\sin\phi\cos\theta \tag{32}$$

$$\sin(\phi+\theta) - \sin(\phi-\theta) = 2\cos\phi\sin\theta \tag{33}$$

In Eqs. (30) to (33) by making the substitution

$$\phi = \tfrac{1}{2}(A+B), \theta = \tfrac{1}{2}(A-B) \tag{34}$$

the following relationships are obtained,

$$\cos A + \cos B = 2\cos\tfrac{1}{2}(A+B)\cos\tfrac{1}{2}(A-B) \tag{35}$$

$$\cos A - \cos B = -2\sin\tfrac{1}{2}(A+B)\sin\tfrac{1}{2}(A-B) \tag{36}$$

$$\sin A + \sin B = 2\sin\tfrac{1}{2}(A+B)\cos\tfrac{1}{2}(A-B) \tag{37}$$

$$\sin A - \sin B = 2\cos\tfrac{1}{2}(A+B)\sin\tfrac{1}{2}(A-B) \tag{38}$$

Inverse trigonometric functions. In the equation $x = \sin y$, y is an angle having x as sine. The symbol $\sin^{-1} x$ (or $\arcsin x$) means the angle having x as sine, or the inverse sine of x. Evidently $\arcsin \frac{1}{2}$ has the values $30° + k360°$ and $150° + k360°$, where k is an integer. A like notation applies to the other functions. For example $\arccos \frac{1}{2}$ has the values $\pm 60° + k360°$. Generally a particular value, called the principal value, is chosen and designated by *Arcsin x*, *Arccos x*, *Arctan x*, etc. Arcsin x and Arctan x are chosen in the range from $-90°$ to $+90°$, and Arccos x and Arccot x are chosen in the range $0°$ to $180°$. Thus Arcsin $\frac{1}{2} = 30°$, Arcsin $(-\frac{1}{2}) = -30°$, Arccos $(\frac{1}{2}) = 60°$, Arccos $(-\frac{1}{2}) = 120°$, Arctan $(-1) = 45°$, and Arccot $(-1) = 135°$.

Solution of rectilinear figures. If a side and another part of a right triangle are known, the triangle can be solved. Consider for example the right triangle ABC of Fig. 7 having $AB = 25.00$ ft, angle $B = 90°$, and angle $A = 60°$. From a table of trigonometric functions, $\tan 60 = 1.732$. Then from Fig. 7,

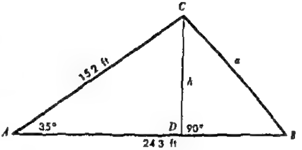


Fig. 8.

$$\frac{BC}{AB} = \frac{BC}{25.00} = \tan 60° = 1.732 \text{ (nearly)}$$

and $\qquad BC = 25.00\,(1.732) = 43.30$ ft

Similarly from Fig. 7,

$$\frac{AB}{AC} = \frac{25.00}{AC} = \cos 60° = 0.5000 \qquad AC = 50.00 \text{ ft}$$

Note that the process of finding an unknown part consists in writing a formula containing the two knowns and the unknown and solving it for the unknown.

A rectilinear figure can often be solved by drawing perpendiculars to sides of the figure to divide it into right triangles and then solving the set of right triangles in order. For example the triangle ABC of Fig. 8 having $AB = 24.3$ ft, $AC = 15.2$ ft, angle $A = 35°$, can be solved by drawing the altitude CD from C to side AB, solving first the right triangle ACD, and then the right triangle CDB.

However various formulas are used for solving rectilinear figures. For convenience, the vertices of a triangle are denoted by capital letters A, B, C, and the respective opposite sides by small letters a, b, c as indicated in Fig. 9.

Law of sines Figure 9 shows a coordinate system with vertex A of a triangle at the origin, side AB on the positive side of the x axis and vertex C above the x axis. By the law of sines

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} \tag{39}$$

Let (x,y) designate the point C in Fig. 9. Then,

$$\frac{y}{b} = \sin A \qquad \frac{y}{a} = \sin B \tag{40}$$

for all angles A and B less than 180°. Dividing the first equation of (40) by the second,

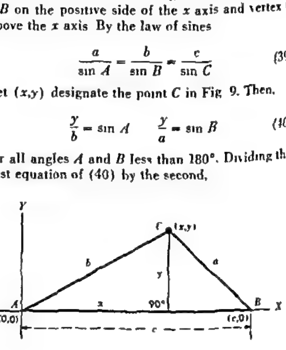


Fig. 9.

$$\frac{a}{b} = \frac{\sin A}{\sin B} \quad \text{or} \quad \frac{a}{\sin A} = \frac{b}{\sin B}$$

A like argument would show that $b/\sin B = c/\sin C$. The law of sines is used in solving triangles in which the given parts are two angles and a side, or two sides and the angle opposite one of them. Thus, if a triangle has $A = 65°$, $B = 40°$, $a = 5.0 \times 10^2$, then $b/\sin 40° = 500/\sin 65°$. Hence $b = 500 \sin 40°/\sin 65° = 350$ nearly. Also from $c \sin 75° = 500/\sin 65°$, $c = 530$ nearly.

Law of cosines The law of cosines may be stated for the general triangle of Fig. 9 as follows:

$$a^2 = b^2 + c^2 - 2bc \cos A \tag{41}$$

From Fig 9,

$$x^2 + y^2 = b^2 \qquad (x - c)^2 + y^2 = a^2$$

Subtracting the first of these equations from the second, member by member, and replacing x in the result by its equal $b \cos A$, Eq (41) is obtained after a slight simplification. Similarly, by interchanging letters in the argument just made,

$$\begin{aligned} b^2 &= a^2 + c^2 - 2ac \cos B \\ c^2 &= a^2 + b^2 - 2ab \cos C \end{aligned} \tag{42}$$

Law of tangents Other formulas relating to triangles can be derived from the law of sines and the law of cosines. The law of tangents may be stated, relative to the triangle of Fig 9, as follows:

$$\frac{\tan \frac{1}{2}(A - B)}{\tan \frac{1}{2}(A + B)} = \frac{a - b}{a + b} \tag{43}$$

From the law of sines

$$\frac{\sin A}{\sin B} = \frac{a}{b} \tag{44}$$

Two equations are formed, one by subtracting 1 from both members of Eq (44), the other by adding 1 to both members, then the first of these equations is divided by the second and simplified to get

$$\frac{\sin A - \sin B}{\sin A + \sin B} = \frac{a - b}{a + b} \tag{45}$$

The numerator and the denominator of the left member of Eq (45) may be replaced by their values from Eqs. (37) and (38), and the result simplified to Eq (43) by using Eq (7) Two other forms of the law of tangents are obtained, one by replacing B by C and b by c in Eq (43) and the other [illegible]

[illegible] for a triangle having $a = 46.0$, $b = 31.0$, $C = 60°$, $(A + B) = 180° - 60° = 120°$, and from Eq (43), $\tan \frac{1}{2}(A - B) = \tan \frac{1}{2}(120°)(46 - 31)/(46 + 31)$ From this equation, $\frac{1}{2}(A - B) = 18.6°$ nearly. Solving this with $\frac{1}{2}(A + B) = 60°$, $A = 78.6°$, $B = 41.4°$ Side c can now be found by the law of sines.

Half-angle formulas. Two half-angle formulas for the general triangle [illegible] Fig. 9 follow:

$$\sin \tfrac{1}{2}A = \sqrt{\frac{(s - b)(s - c)}{bc}} \qquad \cos \tfrac{1}{2}A = \sqrt{\frac{s(s - a)}{bc}} \tag{46}$$

where

$$\begin{aligned} s &= \tfrac{1}{2}(a + b + c) \\ s - a &= \tfrac{1}{2}(-a + b + c) \\ &\cdots \end{aligned} \tag{47}$$

Solving Eq (41) for $\cos A$ and adding 1 to both sides of the resulting equation,

$$\begin{aligned} 1 + \cos A &= 1 + \frac{a^2 + b^2 - c^2}{2ab} = \frac{(a + b)^2 - c^2}{2ab} \\ &= \frac{(a + b + c)(a + b - c)}{2ab} \end{aligned} \tag{48}$$

Substituting in Eq. (48) $2 \cos^2 \frac{1}{2}A$ for $1 + \cos A$, from Eq. (28), and using Eq (47),

$$2 \cos^2 \tfrac{1}{2}A = \frac{2s(2)(s - c)}{2ab} \quad \text{or} \quad \cos \tfrac{1}{2}A = \sqrt{\frac{s(s - c)}{ab}}$$

The formula for $\sin \frac{1}{2}A$ is proved by a like procedure. Also, from Eq (46) by division

$$\begin{aligned} \tan \tfrac{1}{2}A &= \sqrt{\frac{(s - b)(s - c)}{s(s - a)}} \\ &= \frac{1}{s - a} \sqrt{\frac{(s - a)(s - b)(s - c)}{s}} \end{aligned} \tag{49}$$

By interchanging the letters applied to the different parts in the proof of Eq (49), formulas can be derived for $\tan \frac{1}{2}B$ and for $\tan \frac{1}{2}C$. The half-angle formulas can be written

$$\tan \tfrac{1}{2}A = \frac{r}{s - a} \qquad \tan \tfrac{1}{2}B = \frac{r}{s - b} \qquad \tan \tfrac{1}{2}C = \frac{r}{s - c} \tag{50}$$

where

$$r = \sqrt{\frac{(s - a)(s - b)(s - c)}{s}} \tag{51}$$

Formulas (50) are used to solve a triangle for which three sides are known

Using the fact from Eq (25) that $\sin 2(\frac{1}{2}A) = 2 \sin \frac{1}{2}A \cos \frac{1}{2}A$, the fact that the area K of triangle ABC in Fig 9 is $\frac{1}{2}cy = \frac{1}{2}cb \sin A$, and Eq (46) Heron's famous formula for area K of a triangle can be derived,

$$K = \sqrt{s(s - a)(s - b)(s - c)} \tag{52}$$

Also, it is rather easy to derive the formulas for the radii r and R of respective inscribed and circumscribed circles of a triangle

$$r = \sqrt{\frac{(s - a)(s - b)(s - c)}{s}} \qquad R = \frac{abc}{4K} \tag{53}$$

Values of trigonometric functions. If θ is an angle in radians and $n! = 1 \cdot 2 \cdot 3 \cdots n$, then $\sin \theta$ and $\cos \theta$ are defined for all values of θ by the endless series

$$\sin \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots + (-1)^{n+1} \frac{\theta^{2n+1}}{(2n + 1)!} + \cdots \tag{54}$$

$\sin\theta$ and $\cos\theta$, and replacing θ by $\frac{1}{2}\theta$ in the result,

$$\sin\tfrac{1}{2}\theta = \pm\sqrt{(1-\cos\theta)/2} \qquad (27)$$

$$\cos\tfrac{1}{2}\theta = \pm\sqrt{(1+\cos\theta)/2} \qquad (28)$$

Dividing Eq. (27) by Eq (28) member by member,

$$\tan\tfrac{1}{2}\theta = \pm\sqrt{\frac{1-\cos\theta}{1+\cos\theta}} \qquad (29)$$

Conversion formulas. The equations numbered (30) to (33) below are used to transform a product into a sum, and Eqs. (35) to (38) are used to transform a sum into a product. This is generally done for some purpose such as evaluation by logarithms or by a computing machine.

First adding and then subtracting Eqs. (9) and (13), member by member,

$$\cos(\phi+\theta) + \cos(\phi-\theta) = 2\cos\phi\cos\theta \qquad (30)$$

$$\cos(\phi+\theta) - \cos(\phi-\theta) = -2\sin\phi\sin\theta \qquad (31)$$

Similarly from Eqs (10) and (18),

$$\sin(\phi+\theta) + \sin(\phi-\theta) = 2\sin\phi\cos\theta \qquad (32)$$

$$\sin(\phi+\theta) - \sin(\phi-\theta) = 2\cos\phi\sin\theta \qquad (33)$$

In Eqs (30) to (33) by making the substitution

$$\phi = \tfrac{1}{2}(A+B), \theta = \tfrac{1}{2}(A-B) \qquad (34)$$

the following relationships are obtained,

$$\cos A + \cos B = 2\cos\tfrac{1}{2}(A+B)\cos\tfrac{1}{2}(A-B) \qquad (35)$$

$$\cos A - \cos B = -2\sin\tfrac{1}{2}(A+B)\sin\tfrac{1}{2}(A-B) \qquad (36)$$

$$\sin A + \sin B = 2\sin\tfrac{1}{2}(A+B)\cos\tfrac{1}{2}(A-B) \qquad (37)$$

$$\sin A - \sin B = 2\cos\tfrac{1}{2}(A+B)\sin\tfrac{1}{2}(A-B) \qquad (38)$$

Inverse trigonometric functions. In the equation $x = \sin y$, y is an angle having x as sine The symbol $\sin^{-1} x$ (or $\arcsin x$) means the angle having x as sine, or the inverse sine of x. Evidently $\arcsin \frac{1}{2}$ has the values $30° + k360°$ and $150° + k360°$, where k is an integer. A like notation applies to the other functions For example $\arccos \frac{1}{2}$ has the values $\pm 60° + k360°$. Generally a particular value, called the principal value, is chosen and designated by Arcsin x, Arccos x, Arctan x, etc. Arcsin x and Arctan x are chosen in the range from $-90°$ to $+90°$, and Arccos x and Arccot x are chosen in the range $0°$ to $180°$. Thus Arcsin $\frac{1}{2} = 30°$, Arcsin $(-\frac{1}{2}) = -30°$, Arccos $(\frac{1}{2}) = 60°$, Arccos $(-\frac{1}{2}) = 120°$, Arctan $(-1) = 45°$, and Arccot $(-1) = 135°$

Solution of rectilinear figures. If a side and another part of a right triangle are known, the triangle can be solved Consider for example the right triangle ABC of Fig. 7 having $AB = 25.00$ ft, angle $B = 90°$, and angle $A = 60°$. From a table of trigonometric functions, $\tan 60 = 1.732$. Then from Fig 7,

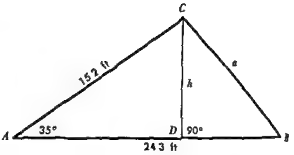


Fig. 8

$$\frac{BC}{AB} = \frac{BC}{25.00} = \tan 60° = 1.732 \text{ (nearly)}$$

and $$BC = 25.00\,(1.732) = 43.30 \text{ ft}$$

Similarly from Fig. 7,

$$\frac{AB}{AC} = \frac{25.00}{AC} = \cos 60° = 0.5000 \qquad AC = 50.00 \text{ ft}$$

Note that the process of finding an unknown part consists in writing a formula containing the two knowns and the unknown and solving it for the unknown

A rectilinear figure can often be solved by drawing perpendiculars to sides of the figure to divide it into right triangles and then solving the set of right triangles in order For example the triangle ABC of Fig. 8 having $AB = 24.3$ ft, $AC = 15.2$ ft, angle $A = 35°$, can be solved by drawing the altitude CD from C to side AB, solving first the right triangle ACD, and then the right triangle CDB

However various formulas are used for solving rectilinear figures. For convenience, the vertices of a triangle are denoted by capital letters A, B, C, and the respective opposite sides by small letters a, b, c as indicated in Fig 9

Law of sines. Figure 9 shows a coordinate system with vertex A of a triangle at the origin, side AB on the positive side of the x axis and vertex C above the x axis. By the law of sines

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} \qquad (39)$$

Let (x,y) designate the point C in Fig 9 Then,

$$\frac{y}{b} = \sin A \qquad \frac{y}{a} = \sin B \qquad (40)$$

for all angles A and B less than $180°$. Dividing the first equation of (40) by the second,

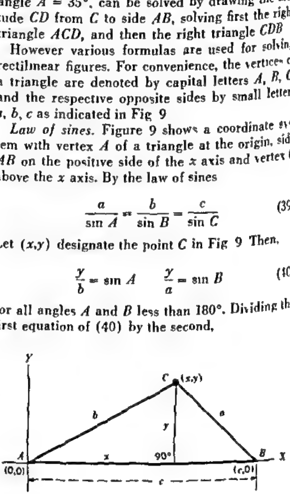


Fig 9

$$\frac{a}{b} = \frac{\sin A}{\sin B} \quad \text{or} \quad \frac{a}{\sin A} = \frac{b}{\sin B}$$

A like argument would show that $b/\sin B = c/\sin C$. The law of sines is used in solving triangles in which the given parts are two angles and a side, or two sides and the angle opposite one of them. Thus, if a triangle has $A = 65°$, $B = 40°$, $a = 5.0 \times 10^2$, then $b/\sin 40° = 500/\sin 65°$. Hence $b = 500 \sin 40°/\sin 65° = 350$ nearly. Also from $c/\sin 75° = 500/\sin 65°$, $c = 530$ nearly.

Law of cosines The law of cosines may be stated for the general triangle of Fig. 9 as follows:

$$a^2 = b^2 + c^2 - 2bc \cos A \tag{41}$$

From Fig 9,

$$x^2 + y^2 = b^2 \qquad (x - c)^2 + y^2 = a^2$$

Subtracting the first of these equations from the second, member by member, and replacing x in the result by its equal $b \cos A$, Eq (41) is obtained after a slight simplification. Similarly, by interchanging letters in the argument just made,

$$\begin{aligned} b^2 &= a^2 + c^2 - 2ac \cos B \\ c^2 &= a^2 + b^2 - 2ab \cos C \end{aligned} \tag{42}$$

Law of tangents Other formulas relating to triangles can be derived from the law of sines and the law of cosines. The law of tangents may be stated, relative to the triangle of Fig 9, as follows.

$$\frac{\tan \frac{1}{2}(A - B)}{\tan \frac{1}{2}(A + B)} = \frac{a - b}{a + b} \tag{43}$$

From the law of sines

$$\frac{\sin A}{\sin B} = \frac{a}{b} \tag{44}$$

Two equations are formed, one by subtracting 1 from both members of Eq (44), the other by adding 1 to both members, then the first of these equations is divided by the second and simplified to get

$$\frac{\sin A - \sin B}{\sin A + \sin B} = \frac{a - b}{a + b} \tag{45}$$

The numerator and the denominator of the left member of Eq (45) may be replaced by their values from Eqs. (37) and (38), and the result simplified to Eq (43) by using Eq (7). Two other forms of the law of tangents are obtained, one by replacing B by C and b by c in Eq (43) and the other by repl[illegible]

[illegible]

are [illegible] for a triangle having $a = 46.0$, $b = 31.0$, $C = 60°$, $(A + B) = 180° - 60° = 120°$, and from Eq (43), $\tan \frac{1}{2}(A - B) = \tan \frac{1}{2}(120°)$ $(46 - 31)/(46 + 31)$. From this equation, $\frac{1}{2}(A - B) = 18.6°$ nearly. Solving this with $\frac{1}{2}(A + B) = 60°$, $A = 78.6°$, $B = 41.4°$. Side c can now be found by the law of sines.

Half-angle formulas Two half-angle formulas for the general triangle of Fig 9 follow:

$$\sin \tfrac{1}{2}A = \sqrt{\frac{(s - b)(s - c)}{bc}} \qquad \cos \tfrac{1}{2}A = \sqrt{\frac{s(s - a)}{bc}} \tag{46}$$

where

$$\begin{aligned} s &= \tfrac{1}{2}(a + b + c) \\ s - a &= \tfrac{1}{2}(-a + b + c) \\ &\cdots \end{aligned} \tag{47}$$

Solving Eq. (41) for $\cos A$ and adding 1 to both sides of the resulting equation,

$$\begin{aligned} 1 + \cos A &= 1 + \frac{a^2 + b^2 - c^2}{2ab} = \frac{(a + b)^2 - c^2}{2ab} \\ &= \frac{(a + b + c)(a + b - c)}{2ab} \end{aligned} \tag{48}$$

Substituting in Eq. (48) $2\cos^2 \frac{1}{2}A$ for $1 + \cos A$, from Eq. (28), and using Eq (47),

$$2\cos^2 \tfrac{1}{2}A = \frac{2s(2)(s - c)}{2ab} \quad \text{or} \quad \cos \tfrac{1}{2}A = \sqrt{\frac{s(s - c)}{ab}}$$

The formula for $\sin \frac{1}{2}A$ is proved by a like procedure. Also, from Eq. (46) by division

$$\begin{aligned} \tan \tfrac{1}{2}A &= \sqrt{\frac{(s - b)(s - c)}{s(s - a)}} \\ &= \frac{1}{s - a}\sqrt{\frac{(s - a)(s - b)(s - c)}{s}} \end{aligned} \tag{49}$$

By interchanging the letters applied to the different parts in the proof of Eq (49), formulas can be derived for $\tan \frac{1}{2}B$ and for $\tan \frac{1}{2}C$. The half-angle formulas can be written

$$\tan \tfrac{1}{2}A = \frac{r}{s - a} \qquad \tan \tfrac{1}{2}B = \frac{r}{s - b} \qquad \tan \tfrac{1}{2}C = \frac{r}{s - c} \tag{50}$$

where

$$r = \sqrt{\frac{(s - a)(s - b)(s - c)}{s}} \tag{51}$$

Formulas (50) are used to solve a triangle for which three sides are known.

Using the fact from Eq (25) that $\sin 2(\frac{1}{2}A) = 2 \sin \frac{1}{2}A \cos \frac{1}{2}A$, the fact that the area K of triangle ABC in Fig. 9 is $\frac{1}{2}cy = \frac{1}{2}cb \sin A$, and Eq (46) Heron's famous formula for area K of a triangle can be derived,

$$K = \sqrt{s(s - a)(s - b)(s - c)} \tag{52}$$

Also, it is rather easy to derive the formulas for the radii r and R of respective inscribed and circumscribed circles of a triangle

$$r = \sqrt{\frac{(s - a)(s - b)(s - c)}{s}} \qquad R = \frac{abc}{4K} \tag{53}$$

Values of trigonometric functions. If θ is an angle in radians and $n! = 1 \cdot 2 \cdot 3 \cdots n$, then $\sin \theta$ and $\cos \theta$ are defined for all values of θ by the endless series

$$\sin \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots + (-1)^{n+1}\frac{\theta^{2n+1}}{(2n + 1)!} + \cdots \tag{54}$$

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots + (-1)^n \frac{\theta^{2n}}{2n!} + \cdots \quad (55)$$

To find a value of the sine or the cosine of an angle, only as many terms of the series are used as are necessary to ensure required accuracy. The error from cutting off the series at any term after the first is less than the numerical value of the first term not used. Thus to find sin 0.1 accurate to six decimal places only the first two terms of the series in Eq. (54) are used to get

$$\sin 0.1 = 0.1 - (0.1)^3/6 = 0.099833 \quad \text{(nearly)}$$

The first unused term was $(0.1)^5/5!$ and $(0.1)^5/5! < 0.0000001$. Hence the error in sin 0.1 due to neglecting all terms of Eq. (54) except the first two is less than 0.0000001. Taking advantage of Eqs. (3) and (19) to (24), it is necessary to compute only values of functions for angles $\frac{1}{4}\pi$ or less. For five decimal place accuracy, 4 terms of Eq. (54) and 4 terms of Eq. (55) are sufficient. Obviously the series in Eqs. (54) and (55) furnish a powerful and convenient method of finding the sine or the cosine of any angle with any required degree of accuracy.

Generalizations of trigonometry. Many interesting generalizations of trigonometry have been made. One of them, because of its importance in advanced mathematics and science, will be considered here. Assume that the complex numbers $a + bi$, where $i^2 = -1$, obey the basic laws of algebra and that the exponential function has been suitably defined. The trigonometric functions can then be introduced by the formulas

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \quad (56)$$

$$\tan\theta = \frac{1}{\cot\theta} = \frac{\sin\theta}{\cos\theta} \qquad \sec\theta = \frac{1}{\cos\theta} \qquad \csc\theta = \frac{1}{\sin\theta} \quad (57)$$

where θ is angle measure in radians. By multiplying the second equation of (56) by i, and adding the result to the first, Eq. (58) is obtained,

$$e^{i\theta} = \cos\theta + i\sin\theta \quad (58)$$

Equating the nth powers of the sides of Eq. (58) De Moivre's theorem is obtained,

$$e^{in\theta} = \cos n\theta + i\sin n\theta = (\cos\theta + i\sin\theta)^n \quad (59)$$

The laws of trigonometry are easily derived from Eqs. (56) and (57). To prove that $\cos(-\theta) = \cos\theta$, replace θ by $-\theta$ in the first equation of (56). This produces no change in the value of the cosine. To prove that $\sin 2\theta = 2\sin\theta\cos\theta$, replace each function by its value from Eq. (56) to get

$$\frac{e^{i2\theta} + e^{-i2\theta}}{2i} = \frac{2(e^{i\theta} - e^{-i\theta})}{2i}\frac{(e^{i\theta} + e^{-i\theta})}{2}$$

The two sides of this equation are equal. Similar proofs apply for all the other laws of trigonometry. However this new trigonometry includes more than the old. For example, in the first equation of (56), replacing θ by $5i$,

$$\cos 5i = \frac{e^{-5} + e^5}{2} = 74.21 \quad \text{(nearly)}$$

When "imaginary angles" are introduced the relation $|\cos\theta| \leqq 1$ no longer applies. This theory involving the imaginary variable θ has many important applications in the theory of electricity. [L.M.K.]

Bibliography: C. T. Holmes, *Trigonometry*, 1951; L. M. Kells, W. F. Kern, and J. R. Bland, *Plane Trigonometry*, 3d ed., 1951; F. W. Sparks and P. K. Rees, *Plane Trigonometry*, 3d ed., 1952.

Trigonometry, spherical

A great circle on a sphere is the intersection of the sphere with a plane through the center. Spherical trigonometry treats for the most part, of spherical triangles having as sides arcs of great circles on a sphere. Two cases of these triangles are highly important. In one case, the vertices are points on the earth and in the other, on celestial bodies such as the sun, planets, and stars. Astronomers, surveyors, and navigators on ships and airplanes apply the formulas of spherical trigonometry to find such values as the time of day, directions of motion, and positions of ships, airplanes, and reference points. Thus spherical trigonometry is basic in astronomy, in certain kinds of surveying, and in navigation. It is also used in mathematics and its applications.

Spherical triangle. Figure 1 represents parts of three planes which form a trihedral angle at the center O of a sphere and which intersect the sphere in a spherical triangle ABC. Each side of the triangle is measured by the corresponding face angle subtended at the center of the sphere and is expressed by the same number of angular units

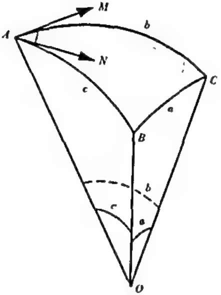


Fig. 1. Spherical triangle.

The tangents AN and AM to the arcs AB and AC at A, being perpendicular to the radius OA, are the sides of a plane angle of dihedral angle M-AO-N. The angle NAM is, by definition, the angle A of spherical triangle ABC. The sides and the angles of spherical triangle ABC are stated in the same angular units. This article will deal only with spherical triangles having sides and angles less than 180°.

Formulas of right spherical triangles. All spherical triangles can be solved by means of ten formulas obtainable from Napier's rules.

Figure 2 shows a right spherical triangle on a sphere of radius 1 unit and center O, and with angle $C = 90°$. Therefore the dihedral angle A-OC-B is a right dihedral angle. Assume that sides a, b, c are all less than 90°. Pass a plane through A perpendicular to edge OB meeting OB in E and OC in D. Then from geometry, angles ADO, ADE, DEO, and AEO are right angles. Hence angle AED = angle B since it is the plane angle of dihedral angle A-OB-C. Figure 2 shows that

$$OE = \cos c \qquad EA = \sin c$$
$$OD = \cos b \qquad DA = \sin b$$

Also from Fig. 2,

$$\cos c = OE = OD \cos a = \cos b \cos a \tag{1}$$

$$\sin b = AD = \frac{AD}{EA} \qquad EA = \sin B \sin c \tag{2}$$

$$\cos B = \frac{ED}{EA} = \frac{ED}{OE} \quad \frac{OE}{EA} = \tan a \cot c \tag{3}$$

$$\sin a = \frac{ED}{OD} = \frac{ED}{DA} \quad \frac{DA}{OD} = \cot B \tan b \tag{4}$$

Similarly by passing a plane through B perpendicular to OA, and deriving

$$\sin a = \sin A \sin c \tag{5}$$
$$\cos A = \tan b \cot c \tag{6}$$
$$\sin b = \tan a \cot A \tag{7}$$

Using the value of cot A obtained from Eq. (7), that of cot B from Eq. (4) and then using Eq. (1),

$$\cot A \cot B = \frac{\sin b}{\tan a} \quad \frac{\sin a}{\tan b} = \cos b \cos a = \cos c \tag{8}$$

Just as Eq. (8) may be derived from Eqs. (1) to (7) the following equations may be derived.

$$\cos B = \cos b \sin A \tag{9}$$
$$\cos A = \cos a \sin B \tag{10}$$

In deriving Eqs. (1) to (10) it was assumed that a, b, and c were less than 90°. If one or more of them were 90°, or greater, at least four parts of triangle ABC would be 90° or greater. Then the triangle would be solved by inspection, and it could be shown by substitution that Eqs. (1) to (10) hold in such cases. Figure 3 represents the orthogonal projection of four triangles composing a hemisphere on the base plane of the hemisphere.

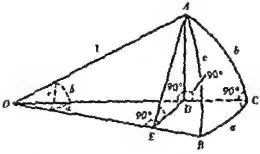


Fig. 2. Right spherical triangle.

In Fig. 3, a, b, c denote sides less than 90°. Of four right triangles composing a hemisphere and each having only one part 90°, one must have three sides each less than 90°. Equations (1) to (10) hold for this spherical triangle, and by substitution, it could then be shown that they hold for the three others. Hence Eqs. (1) to (10) hold for all spherical triangles having parts each less than 180°.

Napier's rules. Equations (1) to (10) can easily be written by two rules. Figure 4 shows the so-called circular parts a, b, Co-A, Co-c, Co-B of a right spherical triangle, the prefix Co- indicates "complement of." Thus Co-B means $(90° - B)$. Each part has two contiguous parts and two noncontiguous parts; speaking of any part as the middle part, the contiguous parts are called adjacent parts, and the noncontiguous parts are called opposite parts. Napier's rules follow.

Rule I. The sine of any middle part is equal to the product of the cosines of the opposite parts.

Rule II. The sine of any middle part is equal to the product of the tangents of the adjacent parts.

The ten equations read by Napier's rules are just the equations (1) to (10). Thus, from Fig. 4, using c as middle part, by rule I,

$$\sin (\text{Co-}c) = \cos c = \cos a \cos b$$

and by rule II,

$$\cos c = \tan (\text{Co-}A) \tan (\text{Co-}B) = \cot A \cot B$$

These are Eqs. (1) and (8) respectively.

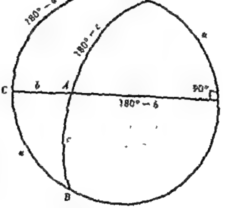


Fig. 3. Orthogonal projection of four triangles.

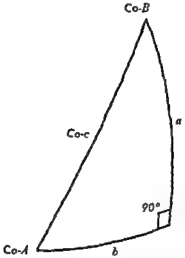


Fig 4. Circular parts of a right spherical triangle.

If two parts of a right spherical triangle are given, the remaining parts can be found since a formula read by Napier's rules applied for two given parts and a desired part has only one unknown and can be solved for that unknown

For example if the given parts are $a = 20°$, $c = 150°$, Napier's rules may be used to obtain

$$\cos 150° = \cos 20° \cos b \qquad \cos B = \tan 20° \cot 150°$$
$$\sin 20° = \sin 150° \sin A$$

Solving, these give $b = 157°$, $B = 129°$, $A = 43°$ nearly. Angle A was taken as 43°, instead of $180° - 43° = 137°$, by rule III, stated below

Whenever the formula used contains the sine of the unknown part, an angle and its supplement are obtained In this case the following rules are used.

Rule III. A nonright angle of a spherical right triangle and the side opposite are both less than 90° or both greater than 90°.

Rule IV. If the legs of a right spherical triangle are both less than 90° or both greater than 90°, the hypotenuse is less than 90°; otherwise it is not less than 90°.

Rule III follows from Eqs. (9) and (10) and rule IV from Eq. (1). There may be two solutions, one solution, or no solution of a right spherical triangle when the given parts are an angle and the side opposite. In this case rules III and IV are used to assign computed parts to the solutions

Polar triangle. If a spherical triangle has parts A, B, C, a, b, c and its polar (*see* POLAR TRIANGLE) parts A', B', C', a', b', c', then

$$A = 180° - a' \qquad B = 180° - b' \qquad C = 180° - c'$$
$$A' = 180° - a \qquad B' = 180° - b \qquad C' = 180° - c \tag{11}$$

Often it is easy to solve the polar triangle of a given triangle. In this case the following rule is used.

Rule V. To solve a triangle Eq. (11) may be used to find the parts of the polar triangle, next the polar triangle is solved, and then the parts of the original triangle are found by using Eq (11).

A quadrantal triangle is one having a side 90°. A polar or a quadrantal triangle for which three parts are known may be solved by using rule V.

Solution of oblique spherical triangles. Spherical triangles are usually classified on the basis of given parts. Six cases are referred to as follows:

1. Given two sides and the included angle.
2. Given two angles and the included side.
3. Given three sides.
4. Given three angles.
5. Given two sides and an angle opposite one of them.
6. Given two angles and a side opposite one of them.

Essentially rule V reduces these six cases to the three, 1, 3, and 5, since the polar triangles associated with cases 2, 4, and 6, are those of cases 1, 3, and 5.

To solve an oblique spherical triangle, divide it into two right triangles which may be solved in succession by Napier's rules. For example take the case in which $a = 118°$, $b = 78°$, $c = 59°$. Figure 5 represents the triangle and the given parts From A an arc perpendicular to side a is drawn meeting it in K Applying Napier's rules one first solves right triangle AKC, then obtains two parts of triangle AKB, and finally solves triangle AKB. From the results the required solution is obtained

The law of sines. Corresponding to the law of sines and the law of cosines in plane trigonometry, there are two basic laws having the same names in spherical trigonometry The law of sines for any spherical triangle having angles A, B, C and respective opposite sides a, b, c is expressed by

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B} = \frac{\sin c}{\sin C} \tag{12}$$

Figure 6 represents any spherical triangle ABC lettered in the conventional way with arc h drawn from C perpendicular to side AB or AB prolonged and meeting it in D Applying Napier's rules first

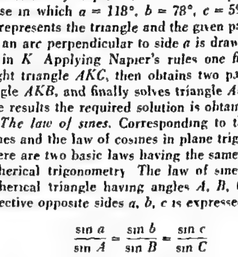


Fig. 5. The triangle and given parts

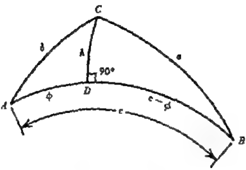


Fig. 6 Spherical triangle.

to right triangle ACD and then to triangle CDB,

$$\sin h = \sin b \sin A \qquad \sin h = \sin a \sin B$$

For these equal values of $\sin h$,

$$\frac{a}{\sin A} = \frac{b}{\sin B}$$

A like procedure proves that $b/\sin B = c/\sin C$.

The law of cosines Assume in Fig 6 that D lies on arc AB, denote arc AD by ϕ and arc DB by $(c - \phi)$ Applying Napier's rules to right triangles ADC and DBC,

$$\cos b = \cos h \cos \phi \tag{13}$$

$$\cos a = \cos h \cos (c - \phi) = \cos h (\cos c \cos \phi + \sin c \sin \phi) \tag{14}$$

$$\cos A = \tan \phi \cot b \tag{15}$$

Dividing Eq (14) by Eq (13) member by member, replacing $\tan \phi$ in the result by $\cos A \tan b$ from Eq (15), and finally multiplying through by $\cos b$,

$$\cos a = \cos b \cos c + \sin b \sin c \cos A \tag{16}$$

It is not necessary to assume that D is on arc AB. If it is assumed that the positive direction of the great circle AB is from A to B, and the negative direction opposite and that ϕ or $(c - \phi)$ may be negative the argument applies generally, unless $b = 90°$ In this case Eq (16) becomes $\cos a = \sin c \cos A$ and this may be proved by applying Napier's rule to the polar of the given triangle By repeating the argument after interchanging the names of parts on Fig 6,

$$\cos b = \cos a \cos c + \sin a \sin c \cos B \tag{17}$$

$$\cos c = \cos a \cos b + \sin a \sin b \cos C \tag{18}$$

If Eqs. (16), (17), and (18) are applied to the polar triangle of ABC by using formulas (11) the law of cosines for angles is obtained.

$$\cos A = -\cos B \cos C + \sin B \sin C \cos a \tag{19}$$

$$\cos B = -\cos A \cos C + \sin A \sin C \cos b \tag{20}$$

$$\cos C = -\cos A \cos B + \sin A \sin B \cos c \tag{21}$$

The law of cosines is very useful in the derivation of other formulas Also, it could be used to solve triangles Evidently if three sides of a triangle are known, Eqs (16), (17) and (18) could be used to find the angles, and if three angles are known Eqs (19), (20), and (21) could be used to find the sides Also if two sides and the included angle are given, one of the equations (16), (17), or (18), would give the third side and then the other two formulas could be used to get the other two angles. Similarly Eqs. (19), (20), and (21) could be used to solve a triangle for which two angles and the included sides are given.

The half-angle formulas. The half-angle formulas of spherical trigonometry follow:

$$\tan \tfrac{1}{2}A = \frac{r}{\sin (s - a)} \qquad \tan \tfrac{1}{2}B = \frac{r}{\sin (s - b)}$$

$$\tan \tfrac{1}{2}C = \frac{r}{\sin (s - c)} \tag{22}$$

where $s = \tfrac{1}{2}(a + b + c)$

$$r = \sqrt{\sin (s - a) \sin (s - b) \sin (s - c)/\sin s} \tag{23}$$

They are derived by a method analogous to that used to get the half-angle formulas of plane trigonometry. Evidently these formulas are used to solve a spherical triangle with three sides given, or, by rule V, one for which three angles are given Other formulas could be obtained by using Eq (11) to apply Eqs. (22) and (23) to the polar triangle and dropping primes. One formula thus obtained involving $S = \tfrac{1}{2}(A + B + C)$ is

$$\tan \tfrac{1}{2}a = \frac{R}{\sin (S - A)}$$

$$R = \sqrt{\frac{\cos (S - A) \cos (S - B) \cos (S - C)}{-\cos S}} \tag{24}$$

Napier's analogies The following formulas, called Napier's analogies are analogous to the law of tangents of plane trigonometry:

$$\frac{\sin \tfrac{1}{2}(A - B)}{\sin \tfrac{1}{2}(A + B)} = \frac{\tan \tfrac{1}{2}(a - b)}{\tan \tfrac{1}{2}c} \tag{25}$$

$$\frac{\cos \tfrac{1}{2}(A - B)}{\cos \tfrac{1}{2}(A + B)} = \frac{\tan \tfrac{1}{2}(a + b)}{\tan \tfrac{1}{2}c} \tag{26}$$

Four others are obtained from these by interchange of letters. Also formulas may be obtained by using Eq (11) to apply Eqs (25) and (26) to the polar triangle. The formulas thus obtained from Eqs. (25) and (26) are

$$\frac{\sin \tfrac{1}{2}(a - b)}{\sin \tfrac{1}{2}(a + b)} = \frac{\tan \tfrac{1}{2}(A - B)}{\cot \tfrac{1}{2}C}$$

$$\frac{\cos \tfrac{1}{2}(a - b)}{\cos \tfrac{1}{2}(a + b)} = \frac{\tan \tfrac{1}{2}(A + B)}{\cot \tfrac{1}{2}C} \tag{27}$$

Observe that if A, B, and C are given, Eqs. (25) and (26) may be used to find a and b The polar formulas and rule V or Eq. (27) may be used to solve a triangle when two sides and the included angle are given

When two sides and the angle opposite one of them (case 5) are given, the law of sines is used to get one angle and then Napier's analogies are used to find the other parts. In this case two solutions, one solution, or no solutions are possibilities. The solution of case 6 is carried out by the same

method applied to the polar triangle. Also the law of sines and Eq. (27) may be used for this case.

[L.M.K.]

Bibliography: Jacques R. Hammond, *Concise Spherical Trigonometry*, 1943; L. M. Kells, W. F. Kern, and J. R. Bland, *Plane and Spherical Trigonometry*, 1951; M. Richardson, *Plane and Spherical Trigonometry*, 1950.

Trihedron

A geometric figure bounded by three noncoplanar rays called edges that emanate from a common point called the vertex, and by the plane sectors called faces that are formed by each pair of edges (Fig 1). A trihedron has three dihedrons formed by pairs of face planes, and three face angles formed by pairs of edges. A plane intersecting the edge of a dihedron cuts it into two trihedrons, whose trihedral angles are measures whose sum is the dihedral angle of the dihedron. Three planes having a common point but not a common line cut space into eight associated trihedrons, of which opposite ones are congruent but not necessarily superposable. (One is the mirror image of the other.) If one of these eight trihedrons has dihedral angles α, β, and γ, and trihedral angle σ, its three neighbors that each share one of its faces will have trihedral angles $\alpha-\sigma$, $\beta-\sigma$, and $\gamma-\sigma$. The sum of the four trihedral angles is 180°, so $2\sigma = \alpha+\beta+\gamma-180°$.

A sphere with center at the vertex of a trihedron cuts the trihedron in a spherical triangle whose angles α, β, γ are the dihedral angles of the trihedron, and whose sides are measured by the face angles of the trihedron (Fig 2). For a discussion of the relations between these angles and sides, *see* TRIGONOMETRY, SPHERICAL; *see also* ANGLE.

[J.S.F.]

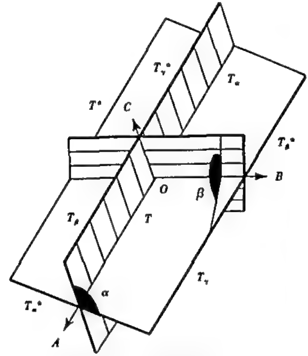


Fig 1 Trihedron and trihedral angles.

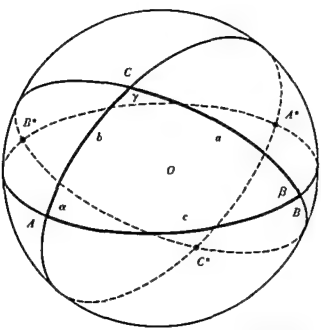
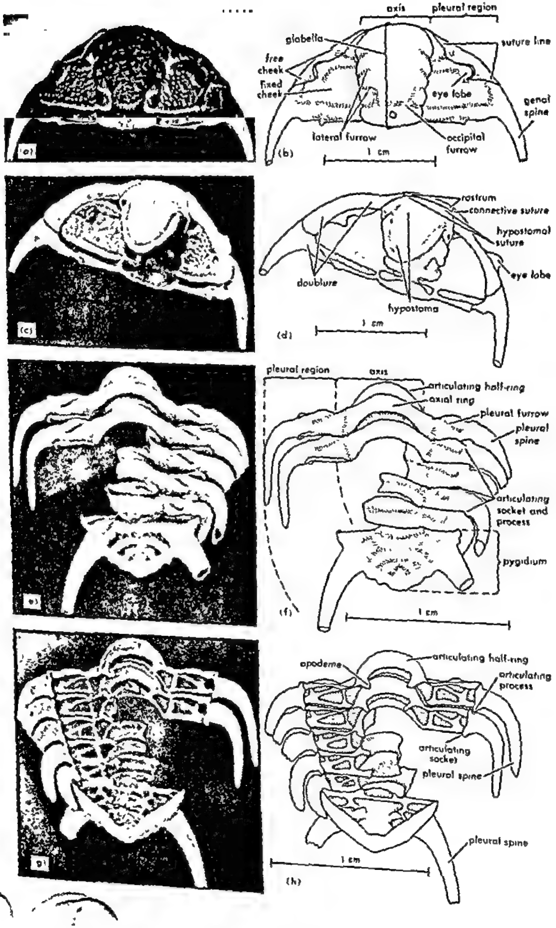


Fig. 2. Spherical triangle ABC formed by intersection of sphere with a trihedron. The center of the sphere is at the vertex of the trihedron.

Trilobita

A class of extinct Paleozoic (Cambrian to Permian) arthropods, occurring with corals, echinoderms, brachiopods (creatures known today only in the seas) and other invertebrates in limestone, shale, and sandstone. They are therefore presumed to have inhabited shallow seas but not fresh waters. The exoskeleton (Fig. 1) covered the dorsal surface and extended a short distance over the ventral side as the doublure It was probably chitinous and strengthened by mineral secretion believed to have been calcite The mineral matter may be preserved in the rock with little change (Fig 4*b*), may be replaced by some other mineral (Fig. 1, 4*c*,*d*), or be dissolved so that an internal (Fig. 4*a*,*f*,*i*) or external mold remains. Exoskeletons in limestone, if replaced by minutely granular quartz, may be freed (Figs. 1, 4*c*) by dissolving the rock in hydrochloric acid which does not attack the quartz. Such specimens reveal both surfaces of the exoskeleton in detail. Trilobites are abundant in many Cambrian sediments (Fig 4*e*), including the earliest, and are used almost exclusively in dating and cor-

Fig 1 Trilobite exoskeletal structures (see facing page) (*a*) *Ceraurus whittingtoni*, Middle Ordovician (Virginia) Silicified exoskeleton of cephalon, upper (dorsal) side (*b*) Outline of original of *a*, showing size and morphological terms (*c*) *Ceraurus whittingtoni*, oblique view of under (ventral) side (*d*) Outline of original of *c*, showing size and morphological terms (*e*) *C. whittingtoni*, five incomplete thoracic segments and pygidium, silicified, dorsal side (*f*) Outline of original of *e*, showing size and morphological terms. (*g*) *C. whittingtoni*, same as *e*, ventral side (*h*) Outline of original of *g*.



axis
pleural region
glabella
free cheek
fixed cheek
suture line
eye lobe
genal spine
lateral furrow
occipital furrow
(a)
(b)
1 cm
rostrum
connective suture
hypostomal suture
eye lobe
doublure
hypostoma
(c)
(d)
1 cm
pleural region
axis
articulating half-ring
axial ring
pleural furrow
pleural spine
articulating socket and process
pygidium
(f)
1 cm
apodeme
articulating half-ring
articulating process
articulating socket
pleural spine
pleural spine
(h)
1 cm

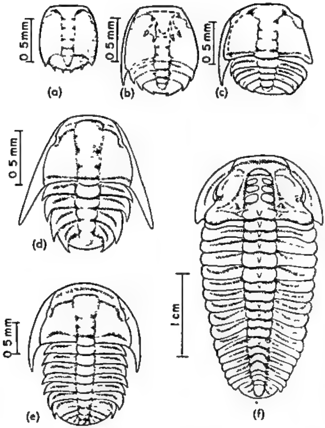


Fig. 2. *Sao hirsuta*, Middle Cambrian, Bohemia (a) Protaspis lacking free cheeks. (b) Protaspis with left free cheek, position of hypostome shown by dashed lines (c) Small meraspid with left free cheek, exoskeleton displaced at joint between cephalon and pygidial portion. (d) Meraspid exoskeleton with one segment in thorax. (e) Meraspid exoskeleton with six segments in thorax. (f) Holaspis (complete exoskeleton) with 17 thoracic segments. (*From H. B Whittington, The ontogeny of trilobites, Biol Revs. Cambridge Phil Soc., 32 437, 1957*)

relating rocks of that system In Ordovician and Silurian rocks, trilobite remains are common, but become fewer in the Devonian, and rare in the Carboniferous and Permian. The beauty of trilobites has long fascinated collectors, and the evolution, taxonomy, and extinction of this large group of more than 1000 genera invite study. An understanding of them is essential in Paleozoic stratigraphy, paleogeography, and animal distribution

Morphology. Trilobita are so called because the body is divided longitudinally into a median axis, lateral to this are the symmetrical pleural regions (Fig. 1*a,e*). An anterior group of fused segments constitute the cephalon (Fig. 1*a,c*), the axial part of which is the glabella. The glabella may be convex and indented by occipital and lateral furrows or may be smooth and ill-defined (Fig. 4*b,g*) The compound eye, situated on the pleural region of the cephalon, or cheek, may exhibit numerous convex facets (Fig. 4*d*) The suture line is an unmineralized line passing over the eye lobe and around the front of the cephalon. On the ventral side of the cephalon (Fig. 1*c*) is a large plate, the hypostome, separated from the doublure by a hypostomal suture; connective sutures isolate a small part of the doublure, the rostrum. The thorax (Fig 1*e,g*) consists of a number of articulated segments each composed of an axial ring and pleurae. Each pleura may be crossed by a pleural furrow and may possess a terminal pleural spine. Articulation between segments is effected by an articulating half-ring that slips under the ring in front, and by small accessory ball and socket joints between the pleurae The pygidium (Fig. 1*e,g*) is composed of fused segments and may bear pleural spines.

Molting and development. Trilobites, like other arthropods, shed their exoskeleton at intervals This molting process was probably helped by a splitting along the suture line, which freed the delicate eye surface and formed an opening through which the animal could crawl. Many fossils are parts of such cast exoskeletons such as the free cheek (part outside the suture), cranidium (rest of cheek and median part of cephalon), rostrum, hypostome, segments, and pygidium. Sometimes these were separated by being washed about on the sea floor (Fig. 3*e*); others are still articulated (Figs 1*e,g*) Molting allowed the animal to increase in

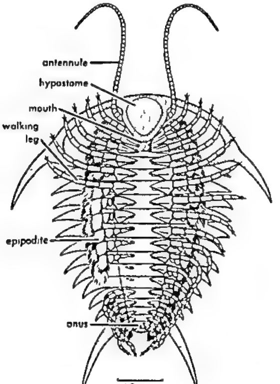


Fig. 3. *Ceraurus pleurexanthemus*, Middle Ordovician, (New York) Restoration of under side, showing appendages. On left side (right side of animal) part or all of appendage removed from thoracic segments 3–9. Ventral membrane stippled Supposed plate behind mouth shown by dotted outline, mouth is concealed by posterior part of hypostome but approximate position is shown. (*From L. Størmer, Studies on trilobite morphology, Norsk Geol Tidsskr., 29 142, 1951*)

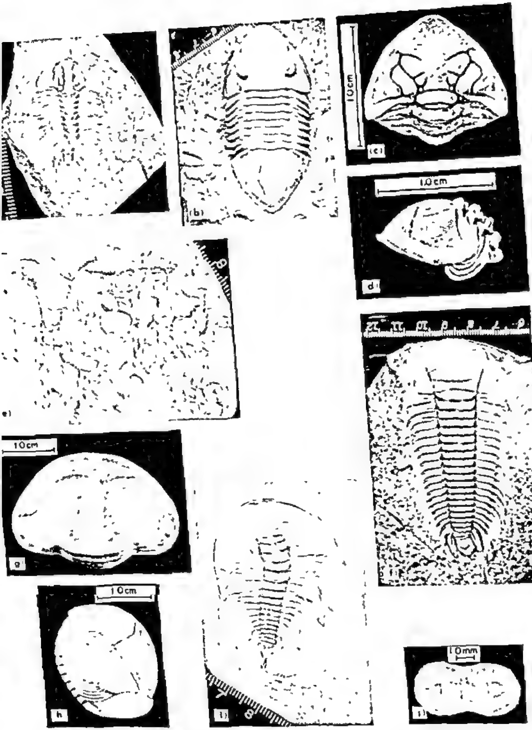


Fig. 4 (a) *Odontopleura ovata*, Silurian, Bohemia Internal mold of exoskeleton, pygidium slightly displaced (b) *Isotelus gigas*, Middle Ordovician, (New York) Exoskeleton partly stripped off internal mold of left cheek and of pygidium (c,d) *Calliops strasburgensis*, Middle Ordovician, (Virginia) Silicified exoskeleton of cephalon and four enrolled thoracic segments (e) Middle Cambrian siltstone, Montana, with scattered parts of many trilobite exoskeletons. (f) *Paradoxides bohemicus*, Middle Cambrian, Bohemia Internal mold of exoskeleton, free cheeks missing (g) *Illaenus lacertus*, Ordovician, Baffin Island Enrolled exoskeleton, upper side of cephalon, dark areas are areas of muscle insertion, (h) right lateral view of *I lacertus* showing eye lobe, sutures, thoracic segments, and pygidium below (i) *Olenellus* sp., Lower Cambrian (British Columbia). Internal mold of exoskeleton. (j) *Ptychagnostus* sp., Middle Cambrian, (Utah) Exoskeleton freed by weathering from shale

size before forming a new exoskeleton, and in some sediments a size series of exoskeletons is preserved, giving a record of development (Fig. 2). The smallest known subhemispherical shield is the protaspis (Fig. *2a,b*). The cephalic portion, with a convex glabella and large eye lobe, is the largest part. As size increases, a joint appears between the cephalic and pygidial portion (Fig *2c*). Segments are formed successively in the terminal part of the pygidial portion, so that the most recently formed is posterior. During the meraspid stage (Fig. *2d,e*), in which the segments increase in size, become fully formed and freely articulating, the thorax is built up to the number of segments characteristic of the species. The thorax gains the full complement of segments when the animal is about 1 cm in length. It is then termed an holaspis, or complete shield, and from then on increase in size occurs without addition of segments to the thorax. Most trilobites reach a length of 2–5 cm but many are longer Specimens of *Isotelus* (Fig *4b*), 40 cm in length, and of *Paradoxides* (Fig. *4j*), about 50 cm in length, are known.

Reconstruction. Rare specimens have appendages preserved as a silvery film on dark shale, or replaced by pyrite. Detailed study has led to the reconstruction of *Ceraurus* (Fig. 3). There was a pair of long, jointed antennules, presumably tactile in function on the cephalon, and four further pairs of appendages. Each pair included a walking leg, from the base of which came a branch carrying fine filaments, the epipodite Each succeeding segment bore similar pairs of appendages None was specialized for grasping, cutting, or grinding food. Thus the trilobite is believed to have walked on the sea floor, or swum above it, and perhaps burrowed in the soft sediment Continuous beating of the epipodites, which perhaps functioned as gills, may have produced currents of water carrying tiny food particles to the mouth. The mouth was probably situated near the back of the hypostome with the stomach and other organs enclosed between the hypostome and glabella, and the alimentary canal extending along the axis to the posterior anus. The musculature in trilobites can only be conjectured, but infolds of the exoskeleton, especially the ventrally-projecting apodemes (Fig. *1g,h*), served for attachments for those muscles at the base of the appendages, and for the muscles that enabled enrollment (Figs *4c,d,g,h*) and extension of the thorax. In trilobites with smooth exoskeletons, darker areas show where muscles were attached (Fig. *4g*).

Evolution. Several kinds of trilobites appear in Cambrian strata (Figs. 2,3,4), and there was a great evolution in later Cambrian times, especially of forms (Figs. *2,4j*) in which the suture line crossed the *posterior cephalic margin* (opisthoparian condition). New groups evolved in the early Ordovician (Figs. *4b,g*) including those in which the suture line crossed the lateral cephalic margin (proparian condition, Figs. *1a,4c*). Other Ordovician trilobites descended, by lines not clearly known, from Cambrian families. No new families arose after the Ordovician, a few continued through the Silurian and Devonian, and only one survived beyond these periods to the end of the Paleozoic.

Variation and distribution. Trilobites exhibit a range in relative size of the cephalon, pygidium, and the number of thoracic segments. The eyes may be absent or present and large; suture lines absent, proparian, or opisthoparian, the furrows numerous and deep or shallow and few, and spines abundant or not developed. The morphological range displayed by the exoskeleton is limited and, so far as we know, appendages were similar in all. In the Cambrian period trilobites outnumbered all other known invertebrates and exceeded them in size Wide geographical distribution of particular genera and even species was possible because the early growth stages floated and drifted. In succeeding periods the shallow seas were inhabited by increasing hordes of other invertebrates, notably Mollusca, including the carnivorous Cephalopoda. Other Arthropoda with pincers were also coming into existence in the middle Paleozoic as well as fishes. The ability to enroll (Fig *4g*) may have served as a protection for trilobites Spininess (Fig. *4a*), thought by some to have been an adaptation to a floating habit, may have made the animal an awkward mouthful for a predator. Competition from better-adapted rivals seems the most obvious reason for the steady decline and eventual extinction of trilobites.

Taxonomy. Trilobite species and genera may be grouped into families fairly readily, but arrangement into higher categories is at present arbitrary and without any agreed framework. Too little is known of the morphology to feel confident of relationships, especially between Cambrian and post-Cambrian forms. Together with the Archaeocyatha, a small group of animals known mainly from the Middle Cambrian Burgess shale, trilobites are regarded today as a separate branch of the Arthropoda. They may have had a common origin with Chelicerata, and be more distantly related to Crustacea. *See* ARCHAEOCYATHA, ARTHROPODA.

[H.B.W.]

Bibliography: R C. Moore (ed), *Treatise on Invertebrate Paleontology*, 1959; J. Piveteau (ed.), *Traité de Paléontologie*, vol. 3, 1953.

Triode, vacuum

A three-element electron tube enclosed in a highly evacuated envelope The tube consists of a cathode, control grid, and anode *See* VACUUM TUBE

The vacuum triode is the fundamental form of all multielement vacuum tubes An understanding of it is basic to a comprehension of all other types of vacuum tubes In the early days of electronics, vacuum triodes were sometimes called pliotrons This had reference primarily to heavy duty tubes for industrial use, and the term is now seldom used.

Control grid. The control grid is the most important and critical electrode in most types of vac

...um tubes. It serves the function of controlling a ...latively large flow of current, and hence power, ...y the application of a relatively small voltage and ...egligible power. The control grid is a screen of ...arallel wires or a square mesh of wires located ...ext to and parallel to the cathode surface and be...tween the cathode and subsequent electrodes It was the invention of the control grid by L. deFor...est that made modern amplifier tubes possible.

The control grid serves to modify the electric ... e ... e are spaces between the wires ... electrons can pass, even though the control grid may be negative relative to the cathode. A typical arrangement of electrodes in a vacuum triode and the resulting potentials are shown in Fig 1 Both the grid and the plate can influence the electric field at the cathode ... ciple ... tined ... Some ... ig 2 This shows the equipotential contours for a single section of an idealized plane-electrode tube This tube has a plane-parallel cathode and plate. with a ... parallel ...

the field changes as the grid is raised from a negative to a positive potential In Fig 2*a* the grid is so negative that the gradient of potential at the cathode is negative, and no current will flow This may also be seen in Fig 3, which shows potential profiles corresponding to the parts of Fig 2 In Fig 2*b* the grid is just at the so-called cutoff potential, that is, its potential is negative and equal to the plate potential divided by the amplification factor Under these conditions the gradient of potential at the cathode is zero, and current may just begin to flow The zero gradient of potential at the cathode is more clearly seen in Fig. 3*b* In Fig 2*c* and Fig 3*c* the grid is negative but high enough in potential so that the gradient of potential opposite the cathode is positive Under these conditions the electrons can flow between the grid wires from cathode to plate. This is the most common condition of operation in such tubes as small-signal amplifiers. Under the conditions given here no current is drawn by the grid. Figure 2*d* and Fig 3*d* show the case when the grid voltage is brought to zero The gradient of potential at the cathode is now strongly positive, and the grid will just begin to draw current. Figure 2*e* and Fig. 3*e* show the case of the positive grid. Here the gradient of potential in front of the cathode is strongly positive, and a large current is drawn Unfortunately, some of this current will be taken by the grid wires This is a condition which is ordinarily avoided in small-signal amplifiers but is often used in high-level radio-frequency amplifiers designed to give high-power amplification However, even here this condition exists only momentarily at the peak of the cycle

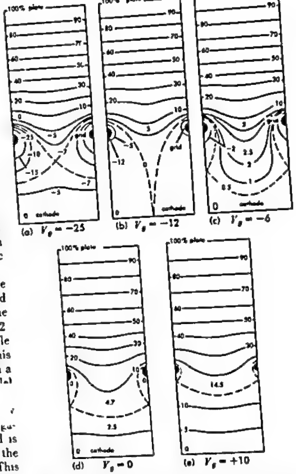


Fig. 2. Equipotential contours in the plane-electrode triode (a) Grid beyond cutoff potential. (b) Grid at cutoff potential (c) Grid negative at half cutoff value (d) Grid at zero potential. (e) Grid positive (From K R Spangenberg, *Vacuum Tubes*, McGraw-Hill, 1948)

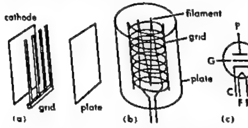


Fig 1 The vacuum triode (a) Basic arrangement of electrodes (b) Typical cylindrical construction including a directly heated cathode (c) Standard circuit symbol (From K R Spangenberg, *Fundamentals of Electron Devices*, McGraw-Hill, 1957)

For comparison, the fields in a cylindrical triode, using a squirrel-cage grid, are shown in Fig 4 It is possible to make such tubes with considerably higher amplification factors than the plane-electrode tubes previously discussed This geometry is

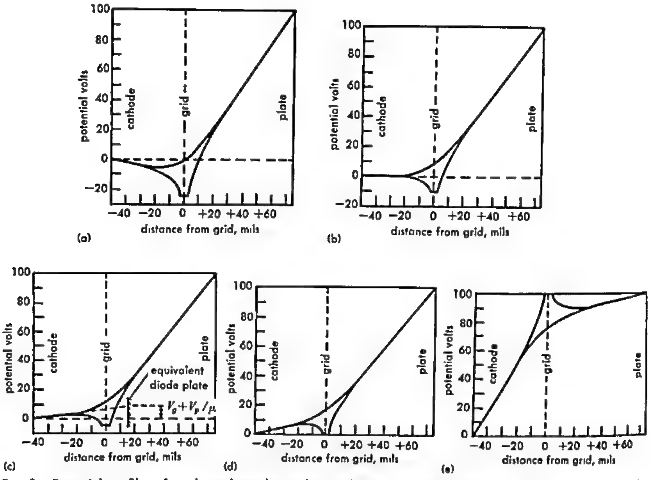


Fig. 3. Potential profiles of a plane-electrode triode. (a) Grid at twice cutoff value of potential. (b) Grid at cutoff potential. (c) Grid at half the cutoff potential. (d) Grid at zero potential. (e) Grid and plate at the same positive potential. *(From K. R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)*

however, only used in certain high-powered transmitting tubes. Most receiving tubes will have a spiral grid which surrounds the cathode.

Triode characteristics. Typical characteristics of the triode are shown in Figs. 5 and 6. Figure 5 shows how the plate current varies with grid voltage for various fixed values of plate voltage. In general, the plate current increases strongly as grid voltage is increased (made less negative). It will also be noticed that there is a large region of useful operation for negative grid voltages where the grid has a large control over the plate current. This is the region used for small-signal amplifiers and is one in which the grid requires virtually no power to drive it. Figure 6 shows the so-called plate characteristics of the tube, that is, the variation of plate current with plate voltage for fixed values of grid voltage. A great deal of information about tube operation can be learned from these characteristics. In Fig. 5 the slope of the characteristic is the so-called mutual conductance g_m of the tube, that is, the ratio of incremental plate current to incremental grid voltage.

$$g_m = \frac{\partial i_b}{\partial e_c}\bigg|_{e_b\text{—constant}}$$

In Fig. 6 the slope of the characteristic is the reciprocal of the plate resistance, r_p. This is the ratio of the incremental plate voltage to the incremental plate current.

$$r_p = \frac{\partial e_b}{\partial i_b}\bigg|_{e_c\text{—constant}}$$

Both of these quantities are fairly constant, although actually the mutual conductance increases approximately as the cube root of the plate current. The plate resistance varies in a reciprocal fashion. The product of the mutual conductance and plate resistance is equal to the amplification factor μ of the tube.

$$\mu = -\frac{\partial e_b}{\partial e_c}\bigg|_{i_b\text{—constant}}$$

$$\mu = r_p g_m$$

In typical triodes the amplification factor will range from 10–100. This is to say that 1 volt change on the control grid is 10 to 100 times more effective in altering the plate current than an equal voltage change on the plate.

Grid current. Also shown in Fig. 6 are curves showing variation of grid current with grid voltage for various values of grid voltage. This grid current flows only when grid voltage is positive relative to the cathode. Grid current decreases as the plate voltage increases, because the electrons move in more nearly straight lines as the plate voltage is increased. As a result, the fraction of the current intercepted by the grid decreases more rapidly than the total current from the cathode. Grid

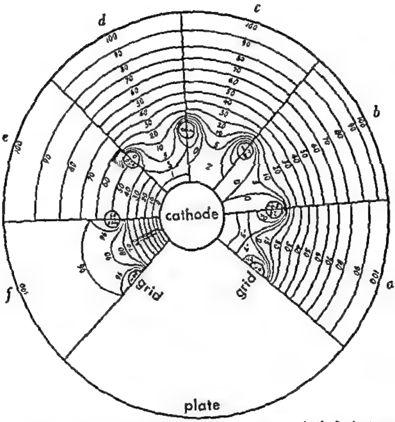


Fig 4 Equipotential contours in the cylindrical-electrode triode (a) Grid beyond cutoff potential (b) Grid at cutoff potential (c) Grid negative but above cutoff potential (d) Grid at zero potential (e) Grid at "natural" potential (f) Grid at positive plate potential (From K R Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)

current is significant only for positive grid operation It represents a driving power needed to operate tubes which is not required when the grid is not driven positive Typical triodes used as radio-frequency amplifiers will have a power amplification in the ratio of 10 20 for radio-frequency operation in which the grid is driven positive at the peak of the cycle.

Applications of triodes. Triodes are extensively used for virtually all operations encountered in electronic devices Their principal functions are, of course, as amplifiers and oscillators, though they may also be used as modulators and detectors as well as wave-shaping and pulse-generating circuits

These same functions may be performed by other tubes but triodes are frequently preferred in some applications As amplifiers, triodes are preferred for audio power output stages because they tend to have less d

also less noisy than multielectrode tubes and are therefore commonly used as the inputs of high-sensitivity receivers and amplifiers Triodes are also preferred in certain special circuits, such as the cathode-follower circuit. This circuit is essentially a 1:1 voltage transformer with, however, a high in- ow output impedance. See

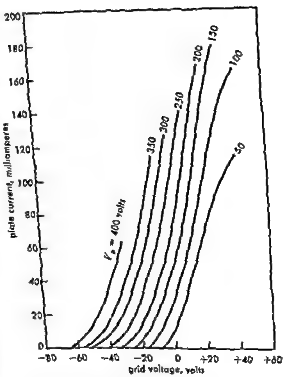


Fig 5 Plate-current versus grid voltage characteristics of a triode (From K R Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)

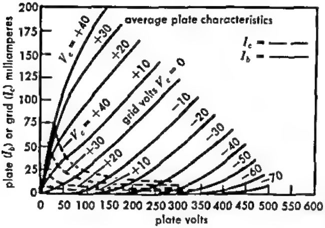
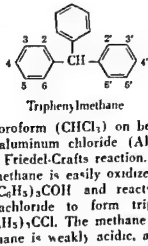


Fig. 6. Plate-current versus plate-voltage characteristics of a triode. *(From K. R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)*

Twin triodes. Twin triodes consist of two complete triode units inside the same vacuum envelope. They are used in a variety of circuits where the operation of two identical tubes is required, most frequently in balanced amplifiers and phase-inverter circuits. They are also used in voltage-regulator circuits and many other places where it is desired to reduce the number of tubes. The use of the twin triodes actually reduces only the number of vacuum envelopes or sockets. The number of effective tubes is still two for each vacuum envelope. *See* ELECTRON TUBE. [K.R.S.]

Triphenylmethane

A colorless, crystalline, aromatic hydrocarbon melting at 92.6°C and boiling at 360°C. Triphenylmethane can be prepared from benzene in one or more steps. The one-step preparation involves the



Triphenylmethane

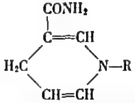
action of chloroform ($CHCl_3$) on benzene in the presence of aluminum chloride ($AlCl_3$), an example of the Friedel-Crafts reaction.

Triphenylmethane is easily oxidized to triphenylcarbinol $(C_6H_5)_3COH$ and reacts with phosphorus pentachloride to form triphenylchloromethane, $(C_6H_5)_3CCl$. The methane hydrogen of triphenylmethane is weakly acidic, and when the hydrocarbon is heated with potassium metal, it evolves hydrogen to yield potassium triphenylmethide, $(C_6H_5)_3CK$.

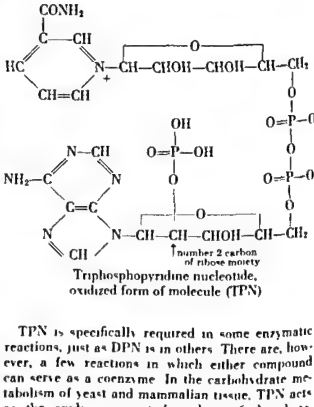
While triphenylmethane has no industrial importance, some of its amino derivatives, for example, 4,4'-bis(dimethylamino)triphenylmethane, form the colorless leuco bases from which the so-called triphenylmethane dyes are prepared. *See* BENZENE; DYE; FRIEDEL-CRAFTS REACTION; POLYNUCLEAR HYDROCARBON. [C.K.B.]

Triphosphopyridine nucleotide (TPN)

A coenzyme and an important component of the enzymatic systems concerned with biological oxidation-reduction systems. It is also known as TPN, triphosphopyridine dinucleotide, coenzyme II, and codehydrogenase II. The compound is similar in structure and function to diphosphopyridine nucleotide (DPN). *See* DIPHOSPHOPYRIDINE NUCLEOTIDE (DPN). It differs structurally from DPN, in having an additional phosphoric acid group esterified at the 2' position of the ribose moiety of the adenylic acid portion. In biological oxidation-reduction reactions the TPN molecule becomes alternately reduced to its hydrogenated form (TPNH) and reoxidized to its initial state.



Triphosphopyridine nucleotide, reduced form of nicotinamide portion of TPNH



Triphosphopyridine nucleotide, oxidized form of molecule (TPN)

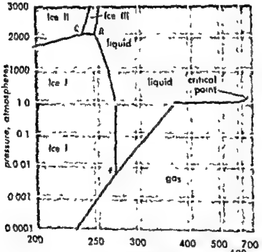
TPN is specifically required in some enzymatic reactions, just as DPN is in others. There are, however, a few reactions in which either compound can serve as a coenzyme. In the carbohydrate metabolism of yeast and mammalian tissue, TPN acts as the oxidizing agent for glucose-6-phosphate, with the enzyme glucose-6-phosphate dehydrogenase:

Glucose-6-phosphate + TPN →
6-phospho-δ-gluconolactone + TPNH + H⁺

TPN is formed, enzymatically, from DPN by phosphorylation with adenosine triphosphate. *See* BIOLOGICAL OXIDATION; CARBOHYDRATE METABOLISM; COENZYME; ENZYME; GLUCOSE-6-PHOSPHATE. [M.D.]

Triple point

A particular temperature and pressure at which 3 different phases of one substance can coexist in equilibrium In common usage, these 3 phases are normally *solid*, *liquid*, and *gas*, although triple points can also occur with 2 solid phases and 1 liquid phase, with 2 solid phases and 1 gas phase, or with 3 solid phases.



Phase diagram for water, showing gas, liquid, and several solid (ice) phases, triple points at *A*, *B*, and *C*. The pressure scale changes at 1 atm from logarithmic scale at low pressure to linear at high pressure.

According to the Gibbs phase rule, a three-phase situation in a one-component system has no degrees of freedom (that is, it is invariant) See EQUILIBRIUM, PHASE Concep...

a unic

in eq

of at least one of the three phases

Triple points are shown in the illustration of part of the phase diagram for water Point *A* is the well known triple point for Ice I (the ordinary low-pressure solid form) + liquid water + water vapor at 0 0099°C (273 16°K) and a pressure of 0.00603 atm (4 68 mm Hg) In 1954, the thermodynamic temperature scale (the absolute or Kelvin scale) was redefined by setting this triple-point temperature for water equal to exactly 273.16°K. Point *B*, at 251 1°K and 2047 atm pressure, is the triple point for liquid water + Ice I + Ice III; and point *C*, at 238 4°K and 2100 atm pressure, is the triple point for Ice I + Ice II + Ice III At least four other triple points are known at higher pressures involving other crystalline forms of ice.

For most substances, the solid liquid-vapor triple point has a pressure less than 1 atm; such substances then have a liquid vapor transition at 1 atm (normal boiling point) However, if this triple point has a pressure above 1 atm, the substance passes directly from solid to vapor at one atmosphere (see SUBLIMATION)

For a two-component system, the invariant point in a phase diagram is a quadruple point at which four phases coexist The three-phase situation is then represented by a line in the 3-dimensional pressure-temperature-composition diagram. *See* BOILING POINT; ICE POINT; MELTING POINT; TRANSITION POINT; VAPOR PRESSURE; WATER. [R.L.S.]

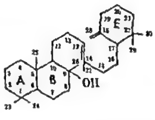
Triterpene

One of a class of chemical compounds having molecular skeletons containing 30 carbon atoms, and theoretically composed of six isoprene units. They are numerous and widely distributed in nature, occurring principally in plant resins and sap, both in the free state and as esters and glycosides. The latter make up an important group known as saponins. In some cases the exact structural formulas have been determined, in many others further studies will be required to define them completely.

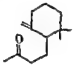
Classification. It is convenient to classify the triterpenes according to the number of rings present Subclassifications are based on the structural similarities between more complex compounds and the simpler monohydric alcohols, for example.

The best known acyclic triterpene is squalene (*see* SQUALENE).

Ambrein is a tricyclic, tertiary alcohol found in ambergris. Permanganate oxidation of ambrein yields some dihydro-γ-ionone

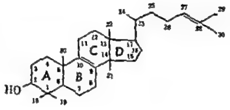


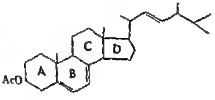
Ambrein



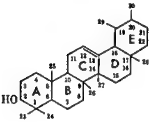
Dihydro-γ-ionone

The tetracyclic triterpenes consist of a small group of naturally occurring compounds called the lanosterol group Recent work, however, indicates a closer relationship to the steroids or sterols

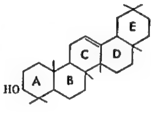




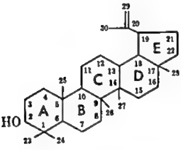
The pentacyclic terpenes are subdivided into three groups, α-amyrin, β-amyrin, and lupeol.



α-Amyrin



β-Amyrin



Lupeol

An example of the α-amyrin group is ursolic acid in which carbon 17 is carboxylated. It is found in the protective, waxlike coating of apples, pears, prunes, and other fruits. Another is β-boswellic acid in which carbon 1 is carboxylated. This acid has been isolated from olibanum and from betulin, the white pigment of birch bark.

An example of the β-amyrin group is glycyrrhetic acid, in which carbon 11 is ketonic and carbon 20 is carboxylated. It occurs in licorice root as a saponin and is conjugated with a hexauronic acid. The ammonium salt of glycyrrhetic acid is the taste principle of licorice.

Therapeutically, glycyrrhetic acid is reported to be the active ingredient of licorice having desoxycortone-like activity. It has been clinically studied as a palliative in Addison's disease.

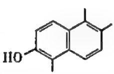
Betulin is a member of the lupeol group in which carbon 28 is hydroxylated. It is present in the bark of the white birch, in other barks, and in lignite.

General reactions. Elucidation of the chemical structure of the triterpenes has been accomplished in many ways. The chemistry of these relatively complex substances involves internal changes in stereoconfigurations and frequently shifting of olefinic bonds. Transformations in molecular structure are possible using $LiAlH_4$

$$R{-}CO{-}Cl \xrightarrow{LiAlH_4} R{-}CH_2{-}OH \xrightarrow{CrO_3} R{-}CHO \rightarrow R{-}CH_3$$

An elegant procedure in establishing the presence of a gem-dimethyl group at the carbon 1 position is shown below:

Valuable information can be obtained by dehydrogenation using selenium or palladized charcoal. However, pentacyclic triterpenes rupture easily into naphthalene fragments. Part of the original triterpene alcohol is preserved in the following naphthol:



In most cases, the hydroxyl groups of the triterpenes are readily acetylated. In both the α- and β-amyrin groups, the olefinic bonds are resistant to hydrogenation. The side-chain olefinic bonds in the lupeol and lanosterol groups undergo the normal reactions of double bonds. *See* Isoprene; Terpene. [E.L.S.]

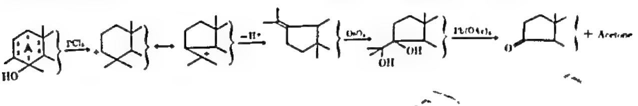
Tritium

The heaviest isotope of the element hydrogen, and the only one which is radioactive. Tritium occurs in very small amounts in nature but is generally prepared artificially by processes known as nuclear transmutations. It is widely used as a tracer in chemical and biological research, and is a component of the so-called thermonuclear or hydrogen bomb. It is commonly represented by the symbol $_1H^3$, indicating that it has an atomic number of 1 and an atomic mass of 3, or by the special symbol T. For information about the other hydrogen isotopes *see* Deuterium; Hydrogen.

The nucleus of the tritium atom, often called a triton and symbolized *t*, consists of a proton and two neutrons. It has a mass of 3.01700 atomic mass units (amu), a nuclear spin of ½, and a magnetic moment of 2.9788 nuclear magnetons. It undergoes radioactive decay by emission of a β-particle to leave a helium nucleus of mass 3. No γ-rays are emitted in this process. The half-life for the decay is 12.26 years. The most energetic of the β-particles emitted by tritium have the comparatively low energy of 18.6 thousand electron volts (kev); β-particles are completely stopped by 7 mm of air or by 0.01 mm of paper or similar material. The average energy of the β-particles is 5.69 kev.

Properties. Both molecular tritium, T_2, and its counterpart hydrogen, H_2, are gases under ordinary conditions. Because of the great difference in mass, many of the properties of tritium differ substantially from those of ordinary hydrogen, as indicated in the table.

Chemically, tritium behaves quite similarly to hydrogen. However, because of its larger mass, many of its reactions take place more slowly than do those of hydrogen. The ratio of reaction rates may be as large as 64:1. These differences in re-



Properties of tritium and hydrogen

	H_2	T_2
Melting point, °C	−259 20	−252 54
Boiling point at 1 atm, °C	−252 77	−248 12
Heat of vaporization, cal/mole	216	333
Heat of sublimation, cal/mole	247	393

activity can give rise to serious errors of interpretation when tritium is used as a tracer for hydrogen

Tritium was first produced by bombarding compounds of deuterium with high-energy deuterons. The nuclear reaction may be written

$$_1H^2 + {}_1H^2 \rightarrow {}_1H^3 + {}_1H^1$$

A number of other nuclear reactions likewise give rise to tritium The most important of these is the absorption of slow neutrons by the lithium isotope of mass 6 according to the equation

$$_3Li^6 + {}_0n^1 \rightarrow {}_1H^3 + {}_2He^4$$

By the use of neutrons from a nuclear reactor, large amounts of tritium may be manufactured in this way, and this is probably the source of the tritium used in the hydrogen bomb

When tritium is bombarded with deuterons of sufficient energy, a nuclear reaction known as fusion occurs, and energy considerably greater than that of the bombarding particle is released. The reaction may be written

$$_1H^3 + {}_1H^2 \rightarrow {}_2He^4 + {}_0n^1 + 18\ \text{Mev}$$

This reaction is one of those which supply the energy of the thermonuclear bomb and may be of importance in the production of energy in controlled thermonuclear reactors

Rain water has been found to contain approximately 1-10 atoms of tritium per 10^{18} atoms of hydrogen This tritium is believed to originate from the bombardment of the nitrogen of the atmosphere with energetic neutrons produced by cosmic rays Because the half-life of tritium is short in comparison with the time required for mixing of the ocean waters the concentration of tritium in the ocean is much lower than that in rain water The total amount of natural tritium on the earth's surface has been estimated at 1800 g. of which about 11 g are in the atmosphere and 13 g in the ground waters Measurement of the nat[illegible]

[illegible] waters to the sea.

Compounds. Very few compounds of pure tritium have been prepared and studied Such compounds would undergo decomposition quite rapidly under the action of the tritium β radiation Tritium oxide, T_2O, has been prepared by oxidation of tritium gas with hot copper oxide or by passing an electric spark through a mixture of tritium and oxygen. Its melting point is 4 49°C, compared with 0° for ordinary water Of much greater importance are compounds, especially organic compounds, in which a small fraction of the hydrogen atoms have been replaced by tritium. Such labeled compounds are useful in studies, by tracer techniques, of the path taken by a particular hydrogen atom or by a whole molecule in a chemical reaction or in a biological process. Tritium-labeled compounds may be prepared by ordinary synthetic chemical methods, such as the hydrogenation of unsaturated compounds with a mixture of hydrogen and tritium, or by the exchange of hydrogen for tritium in the presence of a catalyst such as platinum or a strong acid. The irradiation of a mixture of an organic compound and a lithium salt with neutrons in a nuclear reactor results in the production of energetic tritons, some of which are incorporated into the organic compound Another important labeling procedure consists of the exposure of an organic compound to pure tritium gas in a sealed vessel, the tritium β radiation facilitates the exchange of hydrogen in the compound with tritium in the gas. Because of its weak β-radiation, tritium is not readily measured by the ordinary Geiger-Mueller counter. More efficacious is the introduction of the tritium as a gas inside the counting tube. Alternatively, the ionization of a gas caused by the β-radiation may be measured in an ionization chamber, or the tritium compound may be dissolved in a suitable solvent containing a phosphor, and the light pulses excited by the β-particles counted with a scintillation counter. Tritium gas containing only small amounts of ordinary hydrogen may be analyzed with a mass spectrometer or by measuring the density of the gas See FUSION, NUCLEAR, HEAVY WATER; RADIOACTIVE SPECIES PRODUCED BY COSMIC RAYS; RADIOCHEMISTRY; TRACER, RADIOACTIVE; TRITON.

[L.K.]

Bibliography: W. G. Brown, L. Kaplan, A. R. Van Dyken, and K. E. Wilzbach, Tritium as a tool for industrial and chemical research, *Proc Intern Conf. Peaceful Uses Atomic Energy, Geneva, 1955*, 15:16–23, 1956, M. D. Kamen, *Isotopic Tracers in Biology*, 3d ed., 1957.

Triton

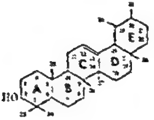
The nucleus of H^3 (tritium), it is the only known radioactive nuclide belonging to hydrogen. The triton is produced in nuclear reactors by neutron absorption in deuterium ($H^2 + n^1 \rightarrow H^3 + \gamma$), and decays by β^- emission to He^3 with a half-life of 12 4 years The spin of the triton is ½, its magnetic moment is 2 9788 nuclear magnetons, and its mass is 3 016449 atomic mass units. Much of the interest in producing H^3 arises from the fact that the fusion reaction $H^3 + H^2 \rightarrow He^4$ releases about 20 Mev of energy Tritons are also used as projectiles in nuclear bombardment experiments. See NUCLEAR REACTION, TRITIUM.

[H.E.D.]

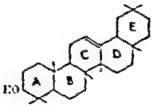
Trochophore

A free-swimming, distinct larval type found in certain marine worms, like polychaetes, nemerteans, and gephyreans, and also in mollusks and Bry-

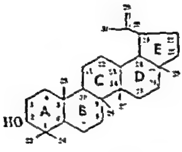
The pentacyclic terpenes are subdivided into three groups, α-amyrin, β-amyrin, and lupeol.



α-Amyrin



β-Amyrin



Lupeol

An example of the α-amyrin group is ursolic acid in which carbon 17 is carboxylated. It is found in the protective, waxlike coating of apples, pears, prunes, and other fruits. Another is β-boswellic acid in which carbon 1 is carboxylated. This acid has been isolated from olibanum and from betulin, the white pigment of birch bark.

An example of the β-amyrin group is glycyrrhetic acid, in which carbon 11 is ketonic and carbon 20 is carboxylated. It occurs in licorice root as a saponin and is conjugated with a hexauronic acid. The ammonium salt of glycyrrhetic acid is the taste principle of licorice.

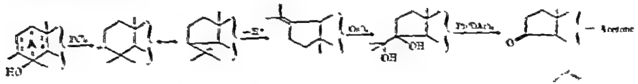
Therapeutically, glycyrrhetic acid is reported to be the active ingredient of licorice having desoxycortone-like activity. It has been clinically studied as a palliative in Addison's disease.

Betulin is a member of the lupeol group in which carbon 28 is hydroxylated. It is present in the bark of the white birch, in other barks, and in lignite.

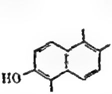
General reactions. Elucidation of the chemical structure of the triterpenes has been accomplished in many ways. The chemistry of these relatively complex substances involves internal changes in stereoconfigurations and frequently shifting of olefinic bonds. Transformations in molecular structure are possible using $LiAlH_4$.

$$R\text{—}CO\text{—}Cl \xrightarrow{LiAlH_4} R\text{—}CH_2\text{—}OH \xrightarrow{CrO_3} R\text{—}CHO \rightarrow R\text{—}CH_3$$

An elegant procedure in establishing the presence of a gem-dimethyl group at the carbon 1 position is shown below:



Valuable information can be obtained by dehydrogenation using selenium or palladized charcoal. However, pentacyclic triterpenes rupture easily into naphthalene fragments. Part of the original triterpene alcohol is preserved in the following naphthol



In most cases, the hydroxyl groups of the triterpenes are readily acetylated. In both the α- and β-amyrin groups, the olefinic bonds are resistant to hydrogenation. The side-chain olefinic bonds in the lupeol and lanosterol groups undergo the normal reactions of double bonds. *See* ISOPRENE; TERPENE. [E.L.S.]

Tritium

The heaviest isotope of the element hydrogen, and the only one which is radioactive. Tritium occurs in very small amounts in nature but is generally prepared artificially by processes known as nuclear transmutations. It is widely used as a tracer in chemical and biological research, and is a component of the so-called thermonuclear or hydrogen bomb. It is commonly represented by the symbol $_1H^3$, indicating that it has an atomic number of 1 and an atomic mass of 3, or by the special symbol T. For information about the other hydrogen isotopes *see* DEUTERIUM; HYDROGEN.

The nucleus of the tritium atom, often called a triton and symbolized *t*, consists of a proton and two neutrons. It has a mass of 3.01700 atomic mass units (amu), a nuclear spin of ½, and a magnetic moment of 2.9788 nuclear magnetons. It undergoes radioactive decay by emission of a β-particle to leave a helium nucleus of mass 3. No γ-rays are emitted in this process. The half-life for the decay is 12.26 years. The most energetic of the β-particles emitted by tritium have the comparatively low energy of 18.6 thousand electron volts (kev); β-particles are completely stopped by 7 mm of air or by 0.01 mm of paper or similar material. The average energy of the β-particles is 5.69 kev.

Properties. Both molecular tritium, T_2, and its counterpart hydrogen, H_2, are gases under ordinary conditions. Because of the great difference in mass, many of the properties of tritium differ substantially from those of ordinary hydrogen, as indicated in the table.

Chemically, tritium behaves quite similarly to hydrogen. However, because of its larger mass, many of its reactions take place more slowly than do those of hydrogen. The ratio of reaction rates may be as large as 64:1. These differences in re-

Properties of tritium and hydrogen

	H_2	T_2
Melting point, °C	−259.20	−252.54
Boiling point at 1 atm, °C	−252.77	−248.12
Heat of vaporization, cal/mole	216	333
Heat of sublimation, cal/mole	247	393

activity can give rise to serious errors of interpretation when tritium is used as a tracer for hydrogen

Tritium was first produced by bombarding compounds of deuterium with high-energy deuterons. The nuclear reaction may be written

$$_1H^2 + {}_1H^2 \rightarrow {}_1H^3 + {}_1H^1$$

A number of other nuclear reactions likewise give rise to tritium The most important of these is the absorption of slow neutrons by the lithium isotope of mass 6 according to the equation

$$_3Li^6 + {}_0n^1 \rightarrow {}_1H^3 + {}_2He^4$$

By the use of neutrons from a nuclear reactor, large amounts of tritium may be manufactured in this way, *and this is probably the source of the* tritium used in the hydrogen bomb.

When tritium is bombarded with deuterons of sufficient energy, a nuclear reaction known as fusion occurs, and energy considerably greater than that of the bombarding particle is released The reaction may be written

$$_1H^3 + {}_1H^2 \rightarrow {}_2He^4 + {}_0n^1 + 18 \text{ Mev}$$

This reaction is one of those which supply the energy of the thermonuclear bomb and may be of importance in the production of energy in controlled thermonuclear reactors

Rain water has been found to contain approximately 1–10 atoms of tritium per 10^{18} atoms of hydrogen This tritium is believed to originate from the bombardment of the nitrogen of the atmosphere with energetic neutrons produced by cosmic rays Because the half life of tritium is short in comparison with the time required for mixing of the ocean waters, the concentration of tritium in the ocean is much lower than that in rain water. The total amount of natural tritium on the earth's surface has been estimated at 1800 g, of which about 11 g are in the atmosphere and 13 g in the ground waters Measurement of the natural tritium content has been used to determine the age of [illegible]

[illegible] very few compounds of pure tritium have been prepared and studied Such compounds would undergo decomposition quite rapidly under the action of the tritium β radiation. Tritium oxide, T_2O, has been prepared by oxidation of tritium gas with hot copper oxide or by passing an electric spark through a mixture of tritium and oxygen Its melting point is 4.49°C, compared with 0° for ordinary water. Of much greater importance are compounds, especially organic compounds, in which a small fraction of the hydrogen atoms have been replaced by tritium. Such labeled compounds are useful in studies, by tracer techniques, of the path taken by a particular hydrogen atom or by a whole molecule in a chemical reaction or in a biological process. Tritium-labeled compounds may be prepared by ordinary synthetic chemical methods, such as the hydrogenation of unsaturated compounds with a mixture of hydrogen and tritium, or by the exchange of hydrogen for tritium in the presence of a catalyst such as platinum or a strong acid. The irradiation of a mixture of an organic compound and a lithium salt with neutrons in a nuclear reactor results in the production of energetic tritons, some of which are incorporated into the organic compound Another important labeling procedure consists of the exposure of an organic compound to pure tritium gas in a sealed vessel; the tritium β-radiation facilitates the exchange of hydrogen in the compound with tritium in the gas. Because of its weak β-radiation, tritium is not readily measured by the ordinary Geiger-Mueller counter More efficacious is the introduction of the tritium as a gas inside the counting tube. Alternatively, the ionization of a gas caused by the β-radiation may be measured in an ionization chamber, or the tritium compound may be dissolved in a suitable solvent containing a phosphor, and the light pulses excited by the β-particles counted with a scintillation counter. Tritium gas containing only small amounts of ordinary hydrogen may be analyzed with a mass spectrometer or by measuring the density of the gas. *See* FUSION, NUCLEAR; HEAVY WATER; RADIOACTIVE SPECIES PRODUCED BY COSMIC RAYS; RADIOCHEMISTRY, TRACER, RADIOACTIVE; TRITON.

[L.K.]

Bibliography W. G Brown, L Kaplan, A. R Van Dyken, and K. E. Wilzbach, Tritium as a tool [illegible] Co., 1951.

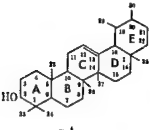
Triton

The nucleus of H^3 (tritium); it is the only known radioactive nuclide belonging to hydrogen The triton is produced in nuclear reactors by neutron absorption in deuterium ($H^2 + n^1 \rightarrow H^3 + \gamma$), and decays by β^- emission to He^3 with a half-life of 12.4 years. The spin of the triton is ½, its magnetic moment is 2.9788 nuclear magnetons, and its mass is 3.016449 atomic mass units. Much of the interest in producing H^3 arises from the fact that the fusion reaction $H^3 + H^3 \rightarrow He^4$ releases about 20 Mev of energy. Tritons are also used as projectiles in nuclear bombardment experiments *See* NUCLEAR REACTION, TRITIUM [H.E.D.]

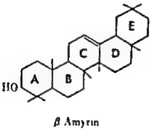
Trochophore

A free-swimming, distinct larval type found in certain marine worms, like polychaetes, nemerteans, and gephyreans, and also in mollusks and Bry-

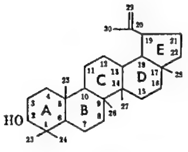
The pentacyclic terpenes are subdivided into three groups, α-amyrin, β-amyrin, and lupeol.



α Amyrin



β Amyrin



Lupeol

An example of the α-amyrin group is ursolic acid in which carbon 17 is carboxylated. It is found in the protective, waxlike coating of apples, pears, prunes, and other fruits. Another is β-boswellic acid in which carbon 1 is carboxylated. This acid has been isolated from olibanum and from betulin, the white pigment of birch bark.

An example of the β-amyrin group is glycyrrhetic acid, in which carbon 11 is ketonic and carbon 20 is carboxylated. It occurs in licorice root as a saponin and is conjugated with a hexauronic acid. The ammonium salt of glycyrrhetic acid is the taste principle of licorice.

Therapeutically, glycyrrhetic acid is reported to be the active ingredient of licorice having desoxycortone-like activity. It has been clinically studied as a palliative in Addison's disease.

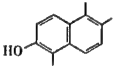
Betulin is a member of the lupeol group in which carbon 28 is hydroxylated. It is present in the bark of the white birch, in other barks, and in lignite.

General reactions. Elucidation of the chemical structure of the triterpenes has been accomplished in many ways. The chemistry of these relatively complex substances involves internal changes in stereoconfigurations and frequently shifting of olefinic bonds. Transformations in molecular structure are possible using $LiAlH_4$.

$$R{-}CO{-}Cl \xrightarrow{LiAlH_4} R{-}CH_2{-}OH \xrightarrow{CrO_3} R{-}CHO \rightarrow R{-}CH_3$$

An elegant procedure in establishing the presence of a gem-dimethyl group at the carbon 1 position is shown below:

Valuable information can be obtained by dehydrogenation using selenium or palladized charcoal. However, pentacyclic triterpenes rupture easily into naphthalene fragments. Part of the original triterpene alcohol is preserved in the following naphthol



In most cases, the hydroxyl groups of the triterpenes are readily acetylated. In both the α- and β-amyrin groups, the olefinic bonds are resistant to hydrogenation. The side-chain olefinic bonds in the lupeol and lanosterol groups undergo the normal reactions of double bonds. *See* ISOPRENE; TERPENE. [E.L.S.]

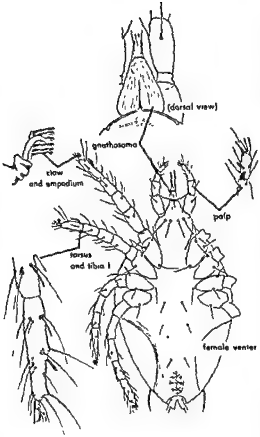
Tritium

The heaviest isotope of the element hydrogen, and the only one which is radioactive. Tritium occurs in very small amounts in nature but is generally prepared artificially by processes known as nuclear transmutations. It is widely used as a tracer in chemical and biological research, and is a component of the so-called thermonuclear or hydrogen bomb. It is commonly represented by the symbol $_1H^3$, indicating that it has an atomic number of 1 and an atomic mass of 3, or by the special symbol T. For information about the other hydrogen isotopes *see* DEUTERIUM; HYDROGEN.

The nucleus of the tritium atom, often called a triton and symbolized *t*, consists of a proton and two neutrons. It has a mass of 3.01700 atomic mass units (amu), a nuclear spin of ½, and a magnetic moment of 2.9788 nuclear magnetons. It undergoes radioactive decay by emission of a β-particle to leave a helium nucleus of mass 3. No γ-rays are emitted in this process. The half-life for the decay is 12.26 years. The most energetic of the β-particles emitted by tritium have the comparatively low energy of 18.6 thousand electron volts (kev); β-particles are completely stopped by 7 mm of air or by 0.01 mm of paper or similar material. The average energy of the β-particles is 5.69 kev.

Properties. Both molecular tritium, T_2, and its counterpart hydrogen, H_2, are gases under ordinary conditions. Because of the great difference in mass, many of the properties of tritium differ substantially from those of ordinary hydrogen, as indicated in the table.

Chemically, tritium behaves quite similarly to hydrogen. However, because of its larger mass, many of its reactions take place more slowly than do those of hydrogen. The ratio of reaction rates may be as large as 64:1. These differences in re-



A trombidiform mite (The Institute of Acarology, University of Maryland)

distinct from those in all other groups The Trombidiformes are probably the most heterogeneous group of mites, both morphologically and ecologically, varying from baglike forms with degenerate legs to the highly evolved, fully developed, parasitic forms There are wormlike forms found in pores of their hosts and flattened types found under scales of lizards, some are parasites in the respiratory tracts of birds, others free living predators of other arthropods, and some are plant-feeders They are also variable in their life histories Some are held within the brood sac of the mother until the siblings or offspring of the same parents have had an opportunity to copulate while others hatch as larvae and pass through a series of molts before becoming sexually mature Economically, this group contains two families of plant feeding mites of great importance to agriculture, the Tetranychidae (spider mites) and the Eriophyidae (bud mites or gall mites) The Tarsonemidae Eupodidae and Tenuipalpidae are of lesser importance The Tetranychidae cause damage by feeding on the leaves and weakening the tree, thus decreasing fruit production, and even causing complete defoliation at times Man, through commerce, has disseminated some of the most important species throughout the world The Eriophyidae, by their feeding, also cause weakening of the trees as well as distortion of fruit and tree growth. A few are vectors of virus diseases of plants. One transmits the streak mosaic virus of wheat in the Middle West; another, peach mosaic virus in the Far West

While some of the Tarsonemini feed on plants, others feed on insects Two species are of particular importance. The hay itch mite, *Pyemotes ventricosus* (Newport), normally lives on insect larvae, but when these are destroyed in the process of harvesting grain, the mites may cause a serious dermatitis on man. The second species, *Acarapis woodi* (Rennie), causes a disease of honey bees in many parts of the world, but so far has not been reported in the United States.

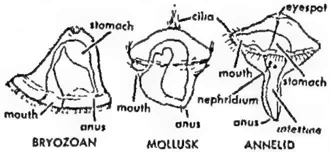
Medically, the Trombiculidae (chiggers, or red bugs) are important because the larval forms, which are parasites of vertebrates, can cause intense irritation to their hosts by their feeding. More seriously, some transmit a rickettsial disease, scrub typhus, to man in the Far East and South Pacific regions. The nymphs and adults of the trombiculids are free-living and prey upon eggs of other arthropods A few families related to the Trombiculidae are predators as nymphs and adults, but the larvae are parasites of arthropods.

The other families are either free-living and are usually considered to be predaceous, or are minor parasites of birds, reptiles, and mammals. Of these, the most important are the Demodicidae or pore mites. *Demodex folliculorum* (Simon) is frequently present in pores of man, but they are seldom noticed and of little medical importance. Species on domesticated animals produce more apparent symptoms and *Demodex canis* (Leydig) is at times fatal for dogs A large and interesting group is the colorful Hydrachnellae which, with few exceptions, are found living in fresh water. A few are parasites of fresh water mussels, and the larvae of others parasitize aquatic insects Adults are predaceous on small aquatic animals *See* PLANT DISEASE; PLANT VIRUS; TYPHUS SCRUB [E.W.B.]

Tropic of Cancer

The parallel of latitude about 23½° (23°45) north of the Equator. The importance of this line lies in the fact that its degree of angle from the Equator is the same as the inclination of the earth's axis from the vertical to the plane of the ecliptic Because of this inclination of the axis and the revolution of the earth in its orbit, the vertical overhead rays of the sun may progress as far north as 23½° At no place north of the Tropic of Cancer will the sun, at noon, be 90° overhead.

On June 21, the summer solstice (Northern Hemisphere), the sun is vertical above the Tropic of Cancer On this same day the sun is 47° above the horizon at noon at the Arctic Circle, and at the Tropic of Capricorn, only 43° above the horizon. The Tropic of Cancer is the northern boundary of the equatorial zone referred to as the tropics, which lies between the Tropic of Cancer and the Tropic of Capricorn *See* GEOGRAPHY, MATHEMATICAL; SOLSTICE [V.H.E.]



Some trochophore larvae: a bryozoan, *Patella*, a mollusk; *Polygordius*, an annelid. (From T. I. Storer and R. L. Usinger, *General Zoology, 3d ed.*, McGraw-Hill, 1957)

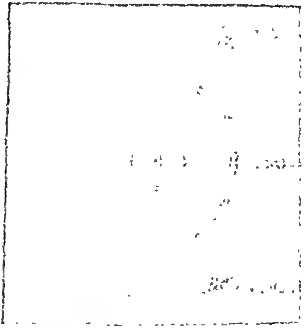
ozoa. Presumably, this larva is indicative of the close evolutionary relationships of these groups. It is provided with a functional gut, paired nephridia, two or more circumferential circlets of cilia (the trochs), a ciliated, sensory apical plate and, sometimes, eyespots. A trochophore is shaped like a pear or like two short cones whose wide bases are fused together. Trochophores metamorphose into adults which look quite different. Originally called trochospheres, their name was changed to trochophore to avoid confusion with *Trochosphaera*, a rotifer, which they superficially resemble [M D RO]

Trogoniformes

An order of birds containing the single pantropical family Trogonidae, the trogons. Although most species are brightly colored, at least in the males trogons are inconspicuous birds, are usually quiet and slow moving, and prefer gloomier parts of tropical forests. The foot structure of trogons is unique among living birds, with the first and second toes, instead of the first and fourth as in many birds, directed backward The relationships of the order are uncertain The best-known species is the brilliantly plumaged quetzal (*Pharomachrus moccino*), the national bird of Guatemala Trogons are hole-nesters, and are primarily fruit eaters The Old World species also eat many insects which they usually capture when in flight *See* AVES. [K.C P]

Trojan planets

Asteroids whose periods of revolution are approximately equal to that of Jupiter, or about 12 years. These bodies move close to one or the other of two positions—60° ahead or 60° behind Jupiter—theoretically found in 1772 by the French mathematician J. L. Lagrange to be stable special solutions of the restricted "three-body" problem (For a discussion of the three-body problem, *see* CELESTIAL MECHANICS.) If an asteroid moves around the Sun at the same mean distance as Jupiter and in the plane of Jupiter's orbit in such a way that the asteroid, the Sun and Jupiter form an equilateral triangle, the motion of the asteroid is stable The first asteroid found to move near one of Lagrange's triangular points was discovered by the German



Lagrangian points and Trojan planets

astronomer Max Wolf at Heidelberg in 1906 and was named Achilles.

Other asteroids of the same family, discovered mainly at Heidelberg, were named after heroes of the Trojan War Fifteen were known in 1959, ten of which, clustered near the first Lagrangian point 60° ahead of Jupiter, form the Greek group, including (588) Achilles, (624) Hektor, (659) Nestor, (911) Agamemnon, and so on; and five, near the second Lagrangian point 60° behind Jupiter, form the Pure Trojan group, including (617) Patroclus, (884) Priamus, and so on.

Actually the Trojan planets do not move exactly in the plane of Jupiter's orbit but sometimes at inclinations in excess of 20° and even 30°, and their mean longitudes differ from that of Jupiter (±60°) sometimes by as much as 10° and even 20°, so that they are occasionally over 10^8 miles from the theoretical Lagrangian points. As a result, their actual motion is very complicated and subject to complex periodic perturbations by Jupiter; furthermore, perturbations by Saturn will eventually remove outlying members of each group, while new members may at long intervals be gained by capture

Because of their great distances from Earth the Trojan planets appear as faint stellarlike objects of twelfth to fourteenth magnitude, although their mean linear diameters are of the order of 70 miles, a relatively large value among minor planets. *See* ASTEROID, PERTURBATION (ASTRONOMY) [G D V]

Trombidiformes

A suborder of the Acarina commonly called the trombidiform mites, more closely related to the Sarcoptiformes than to the other suborders. They are usually distinguished by presence of a respiratory system opening at or near the base of the chelicerae. Other distinguishing characters are to be found in the tarsi, chelicerae, and genitalia. These, although variable within the suborder, are

lated shores, especially those of eastern United States, China, Japan, and eastern Australia.

The main problems concerning tropical storms are determining what starts them, what keeps them going, and where they will move. Although almost all low altitude cyclones drift away from the tropics toward the mid-latitude zones, they move in a great variety of paths. These depend on the arrangement of other wind systems around tropical storms. One of the problems of tropical meteorology is to predict these paths on a daily basis and to estimate seasonally what type of path is likely to develop.

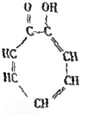
Understanding the formation of cyclones is perhaps the most difficult task. It is known that relative to their outskirts, tropical cyclones are warm in the interior ...

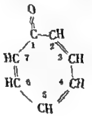
through descent with compression in the innermost center, the eye of tropical storms. Thus the circulation, once well defined, is self-sustaining, but it is difficult to ascertain the circumstances that will lead to its formation. Of the situations showing possibility of hurricane formation, only a few actually produce a hurricane or typhoon storm. Influences from as far away as the middle latitudes play a part, but the conditions that are both necessary and sufficient remain a problem for continuing investigation. [H.RL.]

Bibliography. H. Riehl, *Tropical Meteorology*, 1954.

Tropolone

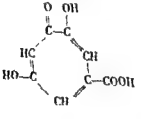
One of a group of organic compounds related to 2-hydroxycycloheptatrienone or tropolone (I). The parent compound is derived from cycloheptatrienone or tropone (II) by substitution of a hydroxyl group for the hydrogen atom in position 2. The hy-


(I)


(II)

drogen atom of the hydroxyl group is replaceable, and the compounds are stronger acids than most phenols, but weaker than the carboxylic acids. Tropone, which ... no acidic ...

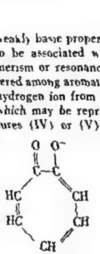
Occurrence. ... system was recognized in 1945 as a structural

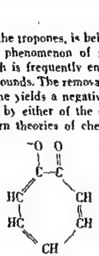

(III)

feature of stipitatic acid (III), a product of the metabolism of certain *Penicillium* mold species. Interest in tropolones developed rapidly as a result of two factors: (1) discovery that a number of natural substances of biological interest possess a tropolone structure and (2) the chemical properties of tropolones, which resemble those of the aromatic compounds.

Among the more important natural tropolone derivatives is colchicine, an alkaloid obtained from meadow saffron (*Colchicum autumnale*). The alkaloid, and in earlier times, the crude plant extract, have been used widely in the treatment of gout. The pronounced effect of colchicine on mitosis has led to its use by modern horticulturists in the production of polyploid plants. Other natural tropolones are the thujaplicins, including hinokitiol (4-isopropyltropolone), found in several species of Cupressaceae ...

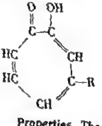
weakly basic properties of the tropones, is believed to be associated with the phenomenon of mesomerism or resonance, which is frequently encountered among aromatic compounds. The removal of a hydrogen ion from tropolone yields a negative ion which may be represented by either of the structures (IV) or (V). Modern theories of chemical

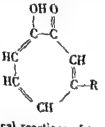

(IV)


(V)

structure predict that molecules which may thus distribute their electrons in different ways possess added stability, the effect of which in this case would be to facilitate loss of the hydrogen ion. Tropone, in behaving as a base, can combine with a proton to form a positive ion whose electrons may be distributed so as to make each of the seven carbon atoms bear a part of the positive charge. The stability of this mesomeric ion serves to make tropone one of the few nonnitrogenous organic bases. See RESONANCE (MOLECULAR STRUCTURE).

In principle, unsymmetrically substituted tropolones could exist as isomers such as (VI) and (VII). However, the separation of such isomeric substances has not been accomplished, apparently because the two forms are easily interconverted.


(VI)


(VII)

Properties. The chemical reactions of tropolone emphasize the stability of the seven ...

Tropic of Capricorn

The parallel of latitude approximately 23½° (23°45) *south of the Equator. It was named for* the constellation Capricornus (the goat), for astronomical reasons which no longer prevail

Because the earth, in its revolution around the sun, has its axis inclined 23½° from the vertical to the plane of the ecliptic, the Tropic of Capricorn marks the southern limit of the zenithal position of the sun. Thus, on December 22 (Southern Hemisphere summer, but northern winter solstice) the sun, at noon, is 90° above the horizon. On this same day, at noon, the sun is 47° above the horizon at the Antarctic Circle, 66½° at the Equator, and 43° at the Tropic of Cancer. Sun rays will just reach the horizon tangentially at the Arctic Circle.

The Tropic of Capricorn is the southern boundary of the equatorial zone referred to as the tropics, which lies between the Tropic of Capricorn and the Tropic of Cancer *See* GEOGRAPHY, MATHEMATICAL; SOLSTICE. [V H E]

Tropical meteorology

The study of atmospheric character in those parts of the world that lie astride the Equator—roughly between 25 or 30° north and south latitude Interest centers mainly on the trade winds and other broad wind systems, on rainfall, on tropical storms, and on these tropical regions as a major heat source for the atmosphere of other parts of the world. Like all meteorology, tropical meteorology involves analysis of the behavior and disturbance patterns of the air near the ground and their correlation with, or control by, high-altitude features of the atmosphere. Such studies form a basis for increasingly useful predictions of precipitation, storms, and other weather conditions of concern in tropical zones. Directly or indirectly, many of these features are an influence on the weather, climate, and related human activities of numerous extratropical areas

Temperature character. In the tropics, where there is no winter *and the temperature is warm to* hot at all times, only the regular diurnal and small-range seasonal cycles bring significant change. Relief from the monotonous heat is most readily obtained by going to the highlands, where air temperatures are cooler with increasing altitude

World heat source. The year-round directness of the sun's rays (*see* GEOGRAPHY, MATHEMATICAL) results in temperature and heat patterns of much greater than tropical significance. It is well known *that the principal* heat source for the atmosphere is situated in the tropics and that heat is exported by ocean currents and atmospheric circulation from there to the higher latitudes. Up to the late 1940s this heat source was assumed to be steady—as if a well-controlled flame had been turned on in a physical experiment *The big changes in middle*-latitude weather were ascribed to variations in the polar regions of the world. The increasing numbers of balloon observations since World War II show, however, that the heat export from the tropics is not constant but is subject to large fluctuations within weeks or months. This interesting discovery suggests that weather on a weekly or monthly basis *in higher latitudes may be affected by these fluctua*tions from the tropics. The computation of transfer of heat, energy, and momentum of the atmosphere from the tropical zone poleward on a short term basis has become a task of tropical meteorology, as has the assessment of the importance of variations in this export upon the weather patterns of the colder belts.

Rainfall and water supply. Though rainfall is high in many parts of the tropics (70–80 in /year and up), abundant water supply is relatively rare The high rate of evaporation caused by the heat greatly reduces the water supply. In addition, the distribution of rainfall over the year is uneven in many areas, especially the monsoon countries, where almost all of the annual rainfall is concentrated in one period of 2–4 months. Hence a delay of some weeks in the onset of the rainy season—a frequent occurrence—will bring serious problems, as will large variation from the normal precipitation of the rainy season—also common.

Causes of tropical precipitation and its variabil ity are studied. For the most part, rainfall yielding copious amounts over wide areas occurs in disturbances of the prevailing trade or monsoon winds in the lower 5 km of the atmosphere. These disturbances have a diameter of about 1000 miles; they take the form of traveling waves, vortices, or lines of sharp cyclonic shear. Their frequency and intensity vary greatly from place to place, and from year to year Relationships of these disturbances with high-altitude wind patterns are studied by means of balloons that measure pressure, temperature, and wind at 8–15 km levels above ground In addition, various meteorological approaches are used to estimate flood rains in mountain areas and thereby aid engineers to develop reservoirs for flood control, irrigation, and water budgeting

Tropical storms. Ordinary disturbances of the weather occur on the average twice weekly in any *area during the rainy season* They produce heavy rain with little wind and little change in barometric pressure Occasionally a central core of intense winds and low barometric pressure develops within these disturbances over the warmest parts of the oceans to form hurricanes and typhoons In these, winds may attain speeds of 100–200 mph in a ring 20–30 miles wide around the center Barometric pressure may drop as much as ten per cent below mean sea level pressure Although their violent core is small, tropical storms as a whole are of about the same size as the cyclones of intermediate and polar latitudes.

Tropical storms are feared because of their destructive power, but they may also be beneficial as relievers of drought. Only limited parts of the tropics are affected by them because they do not occur in all oceans and are seldom encountered within 5° latitude of the Equator. On the other hand, they often move out of the tropics, maintaining their intense wind core, to wreak much e on popu-

Only the lake trout, *Cristovomer namaycush*, and the brook trout, *Salvelinus fontinalis*, are native to the eastern United States. The latter is a char, distinguished by 200 or more small scales in the lateral line, and teeth only in the roof of the mouth. The lake trout is the largest of the family confined to fresh water, attaining a weight of 100 lb in Lake Superior. It occurs in suitable water all across northern North America. The brook trout is one of the smaller trout, averaging only 8–10 in long, although specimens up to 15 lb have been caught. They are native across the Great Lakes drainage, southward into Georgia and northward into Labrador. Like other trout, they have been introduced throughout the world.

The brown trout, *Salmo trutta*, is a European species that is now widely established in the United States and other countries. It tolerates higher temperatures than the brook trout and is somewhat larger, frequently attaining a weight of 12 lb and often growing much larger.

The rainbow trout, *Salmo gairdnerii*, is a western United States species, represented by several subspecies, including a race usually called the steelhead which migrates to the sea. The rainbow trout has been introduced throughout the world, and is the most common of all trout over most of the United States. It has become well established in New Zealand, where it has done well both in size and numbers.

In the western United States these trout are joined by several others, notably the different races of the cutthroat, *Salmo clarkii*. The California golden trout, *Salmo agua bonita*, is generally considered to be the most beautiful of all trout. It is a small mountain-lake species of rather limited range. The Dolly Varden, *Salvelinus malma*, is a Pacific char found from northern California northward and growing to 12 lb. It frequently runs into the sea and is caught in fresh, brackish, and salt water, although essentially it is a fresh-water species. The Dolly Varden is thought by some to be an enemy of the salmon since it frequently eats salmon eggs. Others believe that the Dolly Varden is not a factor in salmon population problems, and much controversy exists over whether it should be protected or considered a pest fish. The Arctic char, *Salvelinus arcturus*, occurs in fresh water only, and is a circumpolar species, ranging southward into southwestern Alaska. There are many other trout species of limited distribution. See CLUPEIFORMES; SALMON. [J D B]

Truck

A wheeled trackless self-propelled vehicle for land transportation of commodities.

Every truck is designed to do a specific job: to haul parcels, miscellaneous commodities, or bulk materials in solid or liquid form, in closely figured amounts over known terrain. Accordingly, there are many models, each representing a combination of components designed to create a unit best suited for the work it is intended to do. A truck is similar to a passenger car in many basic aspects, but its construction is heavier throughout and lower transmission and rear axle ratios are used to cope with hilly terrain (Fig. 1).

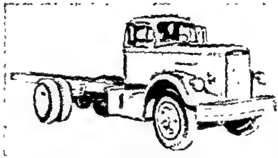


Fig. 1. Conventional truck chassis ready for installation of special body. (*White Motor Co.*)

A truck is rated by its gross vehicle weight (gvw), the combined weight of the vehicle and load. This weight ranges from about 4900 to 65,000 lb. When the gvw is less than 9000 lb, the truck is classed as light weight, as medium weight in the range 9000–16,000 lb, light-heavy in the range 16,000–24,000 lb, and above 24,000 gvw the model is a heavy truck.

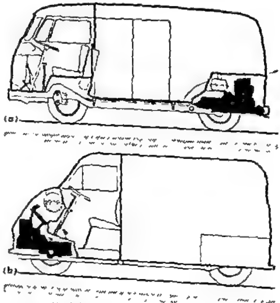


Fig. 2. Usual engine locations arranged for low loading platforms. (a) Engine at rear with rear drive. (b) Engine in front with front-wheel drive. (c) Engine under driver's seat with front-wheel drive.

rated system and its resemblance to compounds of both the aromatic and the aliphatic series of hydrocarbons Thus, the compound undergoes substitution reactions, as do benzene, phenol, and other aromatic compounds, rather than addition reactions, which are more common in the aliphatic series. Nitration, nitrosation, sulfonation, and the diazo-coupling reactions cause substitution at the 5 position in tropolone. Bromination of the sodium or copper salts also occurs at the 5 position, but the 3-bromo compound is formed when tropolone itself is brominated Tropolone is more resistant to both oxidation and reduction than unsaturated aliphatic compounds, but less so than most aromatic compounds. In the presence of a palladium catalyst, hydrogen fails to attack tropolone, but more active catalysts such as platinum will bring about the reduction to cycloheptanol. Although strong oxidizing agents such as permanganate and dichromate cleave the tropolone ring, milder ones do not That even potassium permanganate does not attack it rapidly is demonstrated by the isolation of tropolone from oxidation of cycloheptatriene with this reagent.

The hydroxyl group of tropolone may be replaced in a number of reactions, most of which are analogous to those which replace the hydroxyl group of a carboxylic acid. Alcohols, with acid catalysts, bring about the etherification (replacement of hydroxyl by alkoxyl) of tropolone The resulting ethers, which may be considered as esters, react with amines, alkali, and organometallic reagents just as carboxylic esters do. Thionyl chloride converts tropolones into 2-chlorotropones, which resemble acid chlorides.

The alkali metal salts of tropolones are yellow, and those of the higher metals are deeply colored The complexes with ferric chloride usually are red and may be useful in detecting tropolones The structures of many substituted tropolones have been determined by conversion to substituted benzoic acids under the influence of strong alkali. This reaction involves contraction of the seven-membered ring to a six-membered one and is apparently quite general for the tropolone structure *See* ALICYCLIC HYDROCARBON; ALIPHATIC HYDROCARBON; ALKALOID; AROMATIC HYDROCARBON; PHENOL [R.M ST]

Bibliography: T. Nozoe, The chemistry of natural tropolones and related troponoids, *Festschrift für Arthur Stoll*, 1957; P. L. Pauson, Tropones and tropolones, *Chem. Rev*, 55:9–136, 1955.

Tropopause

Surface of division between troposphere and stratosphere, usually very sharp, but occasionally indefinite and multiple near the jet stream It is highest and coldest in the tropics (17 km and 190°K). In mid-latitudes it lies near 11 km with a temperature of about 210°K; it has a maximum altitude in spring, and the annual thermal wave shows a sharp change of phase there. Day-to-day variations in tropopause temperature are in the opposite sense to variations of troposphere temperature, and the tropopause is high over high-pressure areas and lower over low atmospheric pressure. Near the poles the tropopause is low and comparatively warm. The tropopause was discovered in 1900 by Teisserenc de Bort. *See* ATMOSPHERE; JET STREAM [R.M.C.]

Troposphere

Apart from the lowest 1 or 2 km, the atmosphere up to an average height of about 11 km has a decrease of temperature with height (lapse rate) of about 6.5K°/km independent of latitude and season. This region of uniform lapse rate is the troposphere; its upper boundary is the tropopause, where the lapse rate undergoes a sudden change The nearly uniform lapse rate has often been associated with a pseudoadiabatic lapse rate, such as would accompany intense mixing. The name was given by Teisserenc de Bort. The troposphere is the seat of all important weather phenomena. *See* ATMOSPHERE; TROPOPAUSE. [R M C]

Trout

Any of a number of fishes belonging to the family Salmonidae and native to the colder waters of the Northern Hemisphere. This family also includes the salmon and chars; the latter are usually included under the group designation trout. Although trout are essentially fresh-water fishes, a number of them migrate to the sea where they may grow to large size All of the stream trout spawn in clear, cold, running shallow water, over cleaned gravel Lake trout usually spawn on rocky reefs. All trout eat insects when young, but tend to shift to a diet of fish and other aquatic life as they reach greater size. However, insects remain significant in the diet of most trout throughout life.

Trout. (a) Brook, *Salvelinus fontinalis*, length, 2 yrs, 8–9 in (b) Rainbow, *Salmo gairdnerii*; length, 2 yrs, 8–9 in. (c) Brown, *Salmo trutta*, length, 2 yrs, 8–9 in. (From E L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Truss

A system of structural members lying in a single plane and joined at their ends to form a stable framework. A truss is used like a beam, particularly for bridge and roof construction But because a truss can be made deeper than a beam with solid web and yet not weigh more, it is more economical for long spans and heavy loads

The simplest truss is a triangle composed of three bars with ends pinned together. If small changes in the lengths of the bars are neglected, the relative positions of the joints do not change when loads are applied in the plane of the triangle at the apexes

Types of trusses. Such simple trusses as a triangle, perhaps with the addition of a vertical bar in the middle, are sometimes used to support peaked roofs of houses and other narrow structures. For longer spans, flat roofs, or bridges many triangles are combined to form a truss, as can be seen in the illustrations

In metal trusses, connections may be riveted, bolted, welded, or pinned, in wood trusses, they may be bolted, nailed, or glued Because of long spans, provision must be made to permit movement at one support due to loads and temperature changes: rollers

... the bottom members the lower chord, and the verticals and diagonals, the web members

Framing to be carried by a truss usually is arranged so that it brings loads to bear on the truss at the intersections of a chord and web members. As a result, truss members are subjected only to direct stress—tension or compression—and can be made of le...

sist ...

in te...

pres...

Roof trusses carry the weight of roof deck and framing and wind loads on the upper chord They may also support a ceiling or other loads on the lower chord An example is the Fink truss (Fig 1). On the other hand, bridge trusses may carry loads on either chord Deck trusses support loads on the upper chord (Fig 2); through trusses, on the lower chord (Fig 3)

To maintain stability of truss construction bracing must be used normal to the planes of the trusses Usually framing is inserted between the trusses For roofs, trussed bracing should be placed in the plane of either the top chord or the bottom chord. For bridges, bracing must be inserted in the planes of both top and bottom chords, because of the greater need for stability under heavy moving loads

Computing stresses. Primary stresses in truss members are computed on the assumption that connections at joints are made with frictionless pins With loads applied at joints, each truss member or bar is subjected to pure tension or compression.

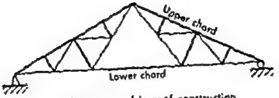


Fig 1. A Fink truss, used in roof construction

Fig. 2. A deck Warren truss, used in bridge construction The load is on the upper chord

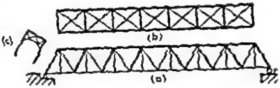


Fig 3. (a) A through Warren truss with verticals Loads are borne on the lower chord (b) Top-chord bracing as seen from above (c) End-on view showing portal bracing.

Since the bars change length under load, the angles of each triangle comprising the truss tend to change. But this change is resisted, since pins are not frictionless, and rivets, bolts, or welds offer restraint Consequently, members bend slightly, the bending moments creating secondary stresses.

At a truss joint, the primary stresses and loads form a coplanar, concurrent force system in equilibrium This system satisfies two conditions: the sums of the horizontal and vertical components both equal zero These equations are used in computing stresses by the "method of joints" In this method, joints with two unknowns are selected in succession and the two equilibrium equations are applied to them to determine the stresses.

A section may be passed through the truss to cut three bars with unknown stresses. These, together with bars with known stresses that are cut and the loads on the part of the structure on either side of the section, constitute a coplanar, nonconcurrent force system in equilibrium. This system satisfies the two previous conditions; but in addition, the sum of the moments of the forces about any axis normal to the plane equals zero. With these three equations the three unknowns can be determined. However, the unknowns also can be found by the "method of moments," in which two of the unknowns are eliminated by taking the moment axis at their point of intersection, and the third is found by equating the sum of the moments to zero The "method of shears" is used to determine one force when the other two unknown forces are both normal to the shearing force, for example, for finding the stresses in the diagonals of parallel-chord Warren, Pratt, and Howe trusses.

By definition there are cab-forward-of-engine (CFE), cab-over-engine (COE), and cab-beside-engine types. The cab may be in fixed position or it may tilt forward to give access to the engine.

Other designations are four-wheel and six-wheel. The four-wheel type drives through the rear wheels only or through all four wheels. The six-wheel type has a tandem rear axle with the drive through one or both axles.

A truck-tractor is a vehicle of short wheelbase for hauling semi-trailers. It carries a swiveling mount, known as a fifth wheel, above the rear axle to support the front end of the semi-trailer. If two axles are used, the drive is through the rear axle, but types with tandem rear axles take the drive through one axle, with one trailing, or through both axles.

A semi-trailer has one or two axles at the rear; the load is carried on these axles and on the fifth wheel of the tractor. The tractor–semi-trailer combination permits the use of longer bodies with greater carrying capacity and better maneuverability than is possible with a conventional truck. Full trailers to be drawn behind semi-trailers have a front axle and one or two rear axles.

The forward positioning of the cab, the short wheelbase of the tractor, and the multiplicity of axles reflect engineering effort to get maximum payloads and operating economy in the face of restriction on over-all length imposed by some states, and regulations limiting the weight carried on a single axle

European manufacturers produce a special class of 1-ton payload trucks which have the engine mounted at the rear with rear-wheel drive, mounted ahead of the front axle with front-wheel drive, or placed under the driver's seat with front-wheel drive (Fig. 2).

Engines are in-line, V type, or pancake and have 4, 6, 8, or 12 cylinders They may operate on gasoline, L-P (liquid petroleum) gas, or diesel fuel The diesel engines operate on a 2-stroke or 4-stroke cycle and are water- or air-cooled Brake horsepower ranges from about 47 at 2800 rpm for a light truck to 356 at 2200 rpm for the heaviest vehicle. Supercharging may be used to develop more horsepower from an engine of given size. *See* SUPERCHARGER

Transmissions have 3, 4, 5, or 7 forward speeds and 1 or 2 reverse speeds. Overdrives are sometimes used By employing a 5-speed transmission in combination with a 2-speed auxiliary transmission, 10 forward speeds are provided. If a truck has a single axle with two speeds, providing two gear ratios within the axle, as many as 12 gear ratios are afforded.

When there is more than one driving axle, an additional gear ratio is provided when needed by an auxiliary transmission known as a transfer case. This is mounted behind the transmission with an output shaft for each of the driven axles.

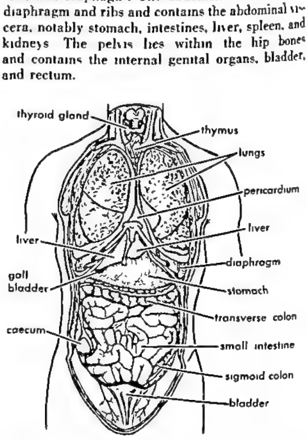
Two types of semi-automatic transmission are used, a straight mechanical and a conventional transmission with hydraulic torque converter. The semi-automatic leaves the driver free to select any gear ratio he chooses and the power shifting spares him the labor of gear changing. Declutching devices, which permit all forward shifts to be made without depressing the clutch pedal, also help the driver. With these the clutch pedal is used only for standing starts.

Alloy steels and aluminum are used in frames Cabs and bodies are framed with steel or aluminum and enclosed with sheet steel, aluminum, or fiber glass reinforced plastics.

Leaf springs are widely used in suspensions Front wheels may or may not be individually sprung. Air springs are employed to some degree when constant frame height and axle articulation, regardless of load, are important considerations *See* AUTOMOTIVE VEHICLE; BULK-HANDLING MACHINES [P.H.S]

Trunk

In man, the main body mass, exclusive of head, neck, and extremities. It is divided into thorax, abdomen, and pelvis The thorax consists of the rib cage area and contains the heart, lungs, great vessels, and esophagus. The abdomen lies below the diaphragm and ribs and contains the abdominal viscera, notably stomach, intestines, liver, spleen, and kidneys The pelvis lies within the hip bones and contains the internal genital organs, bladder, and rectum.



Human trunk. (*From Franz Frohse et al, Atlas of Human Anatomy, rev. ed., Barnes and Noble, 1957*)

The trunk is supported by the vertebral column and back muscles and forms appropriate areas of communication and attachment for neck and limbs.

A similar pattern exists for all vertebrates with modifications typical of certain classes. [E.C.ST.]

d of the undulating membrane or continue as a ee flagellum The organism moves in the direcon of the protruding flagellum. The undulating embrane is a finlike structure composed of two olds of the outer pellicle of the body with the xoneme supporting the outer edge.

Reproduction. Multiplication of the Trypanosoatidae is usually by longitudinal binary fission. For example, in the trypanosomal stage the process s as follows. the kinetoplast divides first and a econd axoneme develops from the new basal granle, as the axoneme increases in length toward the anterior end the nucleus divides; finally when all of the structures have been duplicated the body itself splits longitudinally beginning at the anterior end

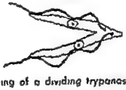


Fig 2 Drawing of a dividing trypanosome

Taxonomy. The six generally recognized genera of the family Trypanosomatidae are *Trypanosoma*, *Leishmania*, *Leptomonas*, *Phytomonas*, *Crithidia* and *Herpetomonas* The first two, *Trypanosoma* and *Leishmania*, are of medical and veterinary importance. Since they are found in the blood of vertebrates, they are referred to as hemoflagellates The other four, *Leptomonas*, *Phytomonas*, *Crithidia*, and *Herpetomonas*, occur only in invertebrates and plants

Genus Trypanosoma. This is the most important genus of the family Trypanosomatidae from a number of standpoints. It contains the largest number of species infecting a wide variety of hosts such as mammals, birds, fishes, amphibians, and reptiles. Although most of the species cause no damage to the hosts, there are several which produce serious diseases in man, his domesticated animals, and wild animals The pathogenic species, prevalent in Africa have been responsible to a great extent for the slow development of civilization in many parts of that continent

Developmental stages Certain species of *Trypanosoma* possess all four developmental stages in their life cycles—trypanosomal, crithidial, leptomonad and leishmanial forms, however, the trypanosomal stage is the most important, being found in the circulating blood of all of the vertebrate hosts Differentiation of the species is based upon the morphology of the blood-form trypanosomes, the particular vertebrates and invertebrates serving as the hosts, the stages present in the hosts, and the course of development and location of the various stages in the hosts

Morphology The trypanosomal stages of the different species differ in size and shape, location of the nucleus, presence and position of the kinetoplast and ... of the undulating membrane. Trypanosomal forms range in length from less than 15 μ to over 80 μ, and may be either slender or broad. In general the nucleus is located near the center, but the kinetoplast may be at the posterior tip or some distance from it. The undulating membrane may be very prominent with many convolutions, or more or less flat and inconspicuous. There may or may not be a free flagellum at the anterior end.

Reproduction Reproduction of the *Trypanosoma* may take place in several stages of the parasite and at different locations in the vertebrate and invertebrate hosts. Typically it is by longitudinal binary fission, but in some instances the division of the cytoplasm is delayed and multiple fission occurs.

Life history. With only rare exceptions, an invertebrate host such as a fly, bug or leech is involved in the transmission of the *Trypanosoma* from vertebrate to vertebrate. In cyclical transmission it is necessary for the parasite to undergo developmental changes in the vector before infective, or metacyclic, trypanosomes occur In the insect, these infective stages, depending upon the species, are located either in the salivary gland or proboscis and are transmitted by the bite as in the case of *T gambiense* and *T rhodesiense*, or are located in the hindgut and are deposited on the surface of the vertebrate when the vector defecates as *T. cruzi*. In the latter case, the infective stage enters the body of the vertebrate host through a break in the skin, sometimes through the hole produced by the bite of the insect. Mechanical transmission can occur with the various species of *Trypanosoma* perhaps without exception, and with some it is the principal means of transmission The trypanosomes thus obtained during a blood meal merely survive on the mouthparts of the insect and are introduced into another vertebrate when the insect feeds again within a few minutes. In *T equiperdum* infections of horses and donkeys, contact transmission occurs during the sexual act and thus no intermediate invertebrate host is required in the life cycle.

Nonpathogenic species. Although the life cycles of the numerous species of *Trypanosoma* may differ in various respects the one for *T. lewisi* serves to illustrate the complexity of the process for a nonpathogenic form. This parasite is worldwide ...

cies

app

Nosopsyllus fasciatus transmits the organism from rat to rat. After the flea ingests the trypanosome stage in the rat's blood, a cyclical development of the parasite takes place in the intestinal tract of the invertebrate host The long blood-form trypanosomes enter the epithelial cells lining the stomach, and reproduce by multiple fission. The modified trypanosomes which emerge from the cells migrate to the hindgut where they attach to the lining by their anterior ends They transform into the crithidial form and divide by binary fission. The infective, short metacyclic stages finally develop and accumulate in the rectum where they pass out with

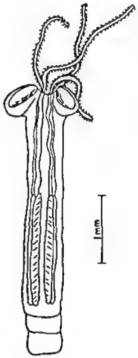
If n is the number of joints in a truss, stresses can be found by the methods of joints, moments, or shears when the number of bars equals $2n - 3$. If a truss is composed of fewer bars, it is unstable; if of more bars, statically indeterminate.

Influence lines are useful in determining the stresses in bridge trusses, because the live load is a moving load. Influence lines can be drawn to show the variation in any function—stress, shear, moment, deflection—as a unit load moves along the truss. *See* STRUCTURAL ANALYSIS. [H L.B.]

Bibliography: *See* STRUCTURAL ANALYSIS.

Trypanorhyncha

An order of tapeworms of the subclass Cestoda also known as the Tetrarhynchoidea. All are parasitic in the intestine of elasmobranch fishes. They are distinguished from all other tapeworm groups by having spiny, eversible proboscides on the head. An elongated head stalk contains the proboscis apparatus made up of a proboscis sheath and a muscular bulb. The head also bears 2 or 4 shallow, weakly muscular suckers (see illustration). Segment anatomy resembles that of Proteocephaloidea, except that the yolk glands are scattered. A complete life history is not known for any trypanorhynchid, although larval forms have been found in the tissues of various marine invertebrates and teleost fishes. *See* CESTODA; *see also* PROTEOCEPHALOIDEA. [C.P R.]



Scolex of *Eutetrarhynchus*.

Trypanosomatidae

A family of Protozoa, order *Protomastigida*, containing flagellated parasites which change their morphology, that is, they exhibit polymorphism during their life cycles. The life cycles of the organisms may involve only an invertebrate host, or an invertebrate and a vertebrate host, or an invertebrate and a plant host. Several distinct morphologic forms are recognized; trypanosomal, crithidial, leptomonad, and leishmanial. Differentiation into genera is dependent upon the host infected as well as the morphologic types involved. None of the stages possess a mouth opening, and nutritive elements are absorbed through the surface of the body; that is, the organisms are saprozoic. The accompanying diagram illustrates the morphologic stages and the hosts of the genera.

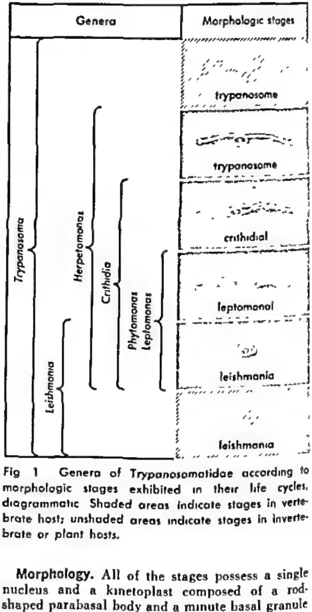


Fig 1 Genera of Trypanosomatidae according to morphologic stages exhibited in their life cycles, diagrammatic Shaded areas indicate stages in vertebrate host; unshaded areas indicate stages in invertebrate or plant hosts.

Morphology. All of the stages possess a single nucleus and a kinetoplast composed of a rod-shaped parabasal body and a minute basal granule called the blepharoplast. In the elongated stages of the parasite, the kinetoplast may be located at the anterior end in the leptomonads, near the center in the crithidials, or at the posterior end in the trypanosomal forms. With the exception of the leishmanial stage, an axoneme or fiberlike structure arises from the basal granule and either extends immediately out of the body as a free flagellum in the leptomonad along the edge of an undulating membrane in the crithidial and trypanosomal forms. The axoneme may terminate at the

entrance into the body through contamination of breaks in the skin or through the conjunctiva of the eye. See HEMIPTERA.

Although *T. cruzi* is apparently present in triatomid bugs and lower animals such as the raccoon, skunk, opossum and armadillo in various parts of the southern half of the United States, the first proved human infection in this country was not reported until 1955

Since 1950, another species, *Trypanosoma rangeli*, has been recognized as a parasite of man In South and Central America this apparently nonpathogenic species must be differentiated from *T cruzi* in examining the intestinal contents of triatomid bugs and the blood of man and other vertebrates

Genus Leishmania. This is the second most important genus, at least from man's standpoint Species of this genus occur as typical leishmanial forms in vertebrate hosts and as leptomonad stages in invertebrate hosts Three species parasitize man but have also been found naturally infecting dogs, cats and perhaps other lower animals. The sandfly, *Phlebotomus*, transmits the parasite from vertebrate to vertebrate. The leishmanial stage after ingestion transforms into the leptomonad form in the gut of the fly and multiplies. The leptomonads are the infective stage for man and are introduced by the bite of the infected fly They enter various cells of the body such as the skin, capillaries spleen and liver, transform to the leishmanial stage, and multiply

The three species of *Leishmania* are morphologically identical Their small, oval bodies, about 5 μ in length, have a relatively large nucleus and kinetoplast, but no flagellum The species are distinguished by their geographical distribution the tissues they infect, and their immune reactions. They all produce serious diseases in man which are difficult to control and treat *L. donovani* infects primarily the internal organs, causing kala-azar or visceral leishmaniasis. *Leishmania tropica* is limited to the surface of the body, producing skin lesions (oriental sore) or cutaneous leishmaniasis *Leishmania brasiliensis* is also limited to the surface of the body but produces skin lesions (espundia, forest yaws) which frequently involve the mucous membranes of the nose, mouth, and pharynx, this is mucocutaneous leishmaniasis See LEISHMANIASIS

Genus Leptomonas. In this genus, the kinetoplast is situated near the anterior end of the elongated body The axoneme arising from the blepharoplast extends directly out of the body as a free flagellum Leptomonads are exclusively parasitic in invertebrates, for example, they occur in the hind gut of the common dog tick *Ctenocephalus canis*. The nonflagellated leishmanial form is the infective stage

Genus Phytomonas. Morphologically similar to *Leptomonas*, phytomonads infect the latex of certain plants for example, milkweed Multiplication of the flagellates may cause degeneration of the infected part of the plant. Hemipterous insects which feed on latex transmit the parasite from plant to plant.

[illegible] Species of this genus are parasit [illegible] is an- [illegible] e are te [illegible] generally a short undulating membrane and a free anterior flagellum. Depending upon the species, the organism may be found in the intestinal tract or body cavity of invertebrates including water bugs and ticks. Leptomonad and leishmanial stages develop, and encysted forms may serve as the infective stage

Genus Herpetomonas. All four stages, trypanosome, crithidial, leptomonas and leishmanial, occur in the life cycle of species in this genus. The several species are exclusively parasitic in invertebrates. They infect the intestinal tracts of various species of flies Transmission is by ingestion of encysted forms. [M M.B.]

Bibliography: C. A. Hoare, *Handbook of Medical Protozoology*, 1950; R R. Kudo, *Protozoology*, 4th ed , 1954

Tryptophan

Tryptophan

Physical constants of the L isomer at 25° C

pK_1 (COOH) 2.38 pK_2 (NH_3^+) 9.39

Isoelectric point 5.89

Optical rotation $[\alpha]_D(H_2O)$ −33.7; $[\alpha]_D$(1 N HCl) +2.8

Solubility (g 100 ml H_2O) 1.14

Absorption spectrum peak at 280 mμ (ultraviolet)

An amino acid considered essential for normal growth of animals. The amino acids are characterized physically by the following. (1) the pK_1, or the dissociation constant of the various titratable groups, (2) the isoelectric point, or pH at which a dipolar ion does not migrate in an electric field; (3) the optical rotation, or the rotation imparted to a beam of plane-polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml; (4) solubility, (5) absorption spectrum, or the wavelength at which maximum absorption occurs See EQUILIBRIUM, IONIC, ISOELECTRIC POINT; OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS.

Tryptophan forms a blue-violet color when treated with glyoxylic acid in the presence of concentrated sulfuric acid (Hopkins-Cole test). Tryptophan is the precursor of several important substances, including the plant growth hormone, indoleacetic acid, the animal hormone, serotonin (5-hydroxytryptamine), the vitamin, nicotinic acid; and certain eye pigments of insects.

The biosynthesis of tryptophan begins with phosphoenolpyruvic acid and D-erythrose-4-phosphate, and proceeds by way of shikimic acid, anthranilic

Species of *Trypanosoma*	Principal vertebrate hosts	Disease	Geographical distribution	Insect vectors	Mode of transmission
T. gambiense	Man and domestic animals	African sleeping sickness	West Equatorial Africa	Tsetse flies (*Glossina*)	Cyclical; insect bite
T. rhodesiense	Probably man and wild animals	African sleeping sickness	East Tropical Africa	Tsetse flies (*Glossina*)	Cyclical, insect bite
T. cruzi	Man, dog, armadillo, opossum, and other animals	Chagas' disease	South and Central America	Kissing bugs (*Triatoma*)	Cyclical; feces of bug
T. brucei	Domestic and wild mammals	Nagana	Tropical Africa	Tsetse flies (*Glossina*)	Cyclical, insect bite
T. vivax	Domestic and wild mammals	Souma	Tropical Africa and South America	Tsetse flies (*Glossina*), Stable flies (*Stomoxys*)	Cyclical (*Glossina*) and mechanical (*Stomoxys*), insect bite
T. equinum	Domestic and wild mammals	Mal de caderas	Tropical and South America	Biting flies (*Tabanus*, *Stomoxys*)	Mechanical, insect bite
T. evansi	Domestic and wild mammals	Surra	Asia, Australia, Madagascar	Biting flies (*Tabanus*, *Stomoxys*)	Mechanical, insect bite
T. hippicum	Domestic mammals, especially horses and mules	Murrina de caderas	Central America	Nonbiting flies (*Musca*)	Mechanical, by flies
T. equiperdum	Horses and donkeys	Dourine	Mediterranean countries	Usually none	Contamination, sexual act

the feces The rat becomes infected through ingestion of the flea's feces contaminating its body or by ingestion of the entire infected flea. For 8 or more days after reaching the blood stream of the rat, the parasites reproduce by multiple fission in the crithidial stage and in other bizarre shapes During this process, the individual organism divides several times without complete fission of the cytoplasm. Finally, the progeny break away from one another and develop into separate trypanosomes At the end of this reproductive period, only typical trypanosomal forms remain In a month or more they are destroyed by the immune response of the host.

Other interesting nonpathogenic species are *T rotatorium* in the frog, which is transmitted by leeches, *T duttoni* in the mouse, *T paddae* in the sparrow; *T granulosum* in the eel; and *T danilewskyi* in goldfish.

Information on a number of the important pathogenic *Trypanosoma* is presented in the accompanying table

Pathogenic species. Some authorities believe that the species which cause African sleeping sickness, *T. gambiense* and *T. rhodesiense*, represent the wild animal species, *T brucei*, which have become adapted to the human body; ordinarily the serum of man is trypanocidal for *T. brucei*. In any event, they are very similar morphologically and exhibit comparable development in the tsetse fly, *Glossina* After the African sleeping sickness parasites enter the human body during the bite of the tsetse fly, these parasites first multiply in the lymph and blood Later they may invade the nervous system Only the trypanosome stage occurs in man In the tsetse flies *Glossina palpalis* and *G morsitans*, multiplication takes place first in the midgut while in the trypanosome stage. Later the parasites migrate to the salivary glands where they transform into crithidial forms and multiply. Eventually metacyclic trypanosomes develop in this site. *See* DIPTERA; SLEEPING SICKNESS, AFRICAN.

Trypanosoma cruzi, the cause of Chagas' disease, presents a somewhat different cycle in man and triatomid bugs These rather large arthropods, about 1 in in length, frequently feed around the face of the sleeping person and therefore are commonly referred to as the kissing or barber bug In the body of man, the trypanosomal stages circulating in the blood stream do not divide All multiplication takes place in tissue cells The parasites enter various tissues including the heart muscles, transform into leishmanial stages, and divide by binary fission until a cluster of parasites fills each cell Before breaking out of the cells into the blood stream, they return to the trypanosomal stage The parasite may reenter other cells and multiply again The trypanosome, after its ingestion by the kissing bug, multiplies first in the midgut and later in the rectum Multiplication in the bug is primarily during the crithidial stage. Metacyclic trypanosomes develop in the rectum and pass out of the bug when it defecates. The infective stages therefore are not inoculated into the vertebrate but gain

ance into the body through contamination of iks in the skin or through the conjunctiva of the *See* HEMIPTERA

lthough *T. cruzi is apparently* present in trimid bugs and lower animals such as the raccoon, nk, opossum and armadillo in various parts of southern half of the United States, the first ted human infection in this country was not reted until 1955

ince 1950, another species, *Trypanosoma rangeli*, has been recognized as a parasite of man In th and Central America this apparently nonhogenic species must be differentiated from *cruzi* in examining the intestinal contents of atomid bugs and the blood of man and other verbrates.

Genus Leishmania. This is the second most imrtant genus, at least from man's standpoint. Species of this genus occur as typical leishmanial rms in vertebrate hosts and ... ins ... ut h...

ls, and perhaps other lower animals The sandfly, *Phlebotomus*, transmits the parasite from vertebrate to vertebrate The leishmanial stage after ingestion transforms into the leptomonad form in the gut of the fly and multiplies The leptomonads are he infective stage for man and are introduced by he bite of the infected fl...

...

The three species of *Leishmania* are morphologically identical. Their small, oval bodies, about 5 μ in length, have a relatively large nucleus and kinetoplast, but *no* flagellum The species are distinguished by their geographical distribution, the tissues they infect, and their immune reactions. They all produce serious diseases in man which are difficult to control and treat. *L. donovani* infects primarily the internal organs causing kala-azar or visceral leishmaniasis *Leishmania tropica* is limited to the surface of the body, producing skin lesions (oriental sore) or cutaneous leishmaniasis. *Leishmania braziliensis* is also limited to the surface of the body, but produces skin lesions (espundia (forest yaws) which frequently involve the mucous membranes of the nose, mouth, and pharynx, this is mucocutaneous leishmaniasis. *See* LEISHMANIASIS

Genus Leptomonas. In this genus, the kinetoplast is situated near the anterior end of the elongated body The axoneme arising from the blepharoplast extends directly out of the body as a free flagellum Leptomonads are exclusively parasitic in invertebrates, for example, they occur in the hindgut of the common dog tick *Ctenocephalus canis* The nonflagellated leishmanial form is the infective stage.

Genus Phytomonas. Morphologically similar to *Leptomonas*, phytomonads infect the latex of certain plants, for example milkweed Multiplication of the flagellates may cause degeneration of the infected part of the plant. Hemipterous insects which feed on latex transmit the parasite from plant to plant.

Genus Crithidia. Species of this genus are parasitic in various arthropods. The kinetoplast is anterior to the centrally placed nucleus There are generally a short undulating membrane and a free anterior flagellum Depending upon the species, the organism may be found in the intestinal tract or body cavity of invertebrates including water bugs and ticks Leptomonad and leishmanial stages develop, and encysted forms may serve as the infective stage.

Genus Herpetomonas. All four stages, trypanosome, crithidial, leptomonas and leishmanial, occur in the life cycle of species in this genus The several species are exclusively parasitic in invertebrates. They infect the intestinal tracts of various species of flies, Transmission is by ingestion of encysted forms. [M M B.]

Bibliography: C. A Hoare, *Handbook of Medical Protozoology*, 1950; R. R. Kudo, *Protozoology*, 4th ed., 1954

Tryptophan

Physical constants of the L isomer at 25° C
pK_1 (COOH): 2.38 pK_2 (NH_3^+): 9.39
Isoelectric point: 5.89
Optical rotation: $[\alpha]_D$ (H_2O): −31.7 $[\alpha]_D$ (5 N HCl): +2.8
Solubility (g/100 ml H_2O): 1.14
Absorption spectrum: peak at 280 mμ (ultraviolet)

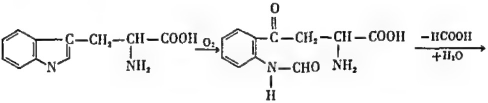
An amino acid considered essential for normal growth of animals. The amino acids are characterized physically by the following: (1) the pK_1, or the dissociation constant of the various titratable groups, (2) the isoelectric point, or pH at which a dipolar ion does not migrate in an electric field, (3) the optical rotation, or the rotation imparted to a beam of plane-polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml; (4) solubility; (5) absorption spectrum, or the wavelength at which maximum absorption occurs. *See* EQUILIBRIUM, IONIC; ISOELECTRIC POINT; OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS

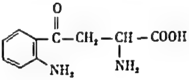
Tryptophan forms a blue-violet color when treated with glyoxylic acid in the presence of concentrated sulfuric acid (Hopkins-Cole test). Tryptophan is the precursor of several important substances, including the plant growth hormone, indoleacetic acid, the animal hormone ...

acid, and indoleglycerol phosphate (*see* AMINO ACIDS).

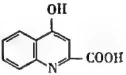
During metabolic degradation, several pathways exist, including the following:

1. Peroxidation of the heterocyclic ring to formylkynurenine, which is deformylated to kynurenine:

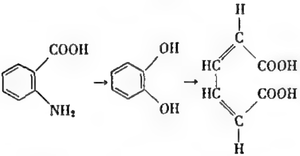




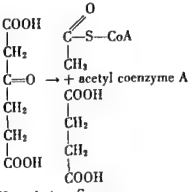
Kynurenine then is metabolized further by any of three alternative routes: (*a*) deamination to the α-keto acid, which spontaneously cyclizes to kynurenic acid:



(*b*) cleavage to alanine and anthranilic acid, anthranilic acid is oxidized, in certain bacteria, to succinate and acetyl coenzyme A.



Anthranilic acid | Catechol | *cis,cis*-Muconic acid



β-Ketoadipic acid | Succinic acid

(*c*) Oxidation to 3-hydroxykynurenine, which is converted to nicotinic acid via 3-hydroxyanthranilic acid.

2. Cleavage of the side chain by the enzyme tryptophanase, forming indole, pyruvic acid, and ammonia.

3. Decarboxylation to tryptamine, and oxidation of the latter to indoleacetic acid.

4 Oxidation to 5-hydroxytryptophan, followed by formation of 5-hydroxytryptamine and 5-hydroxyindoleacetic acid. [E A AD]

Tsunami

A long surface gravity wave (tidal wave) formed by an impulsive dislocation of the sea floor. The dislocation is usually associated with a shallow focus (<30 km) earthquake of intensity greater than seven on the Gutenberg-Richter scale. However, not all such earthquakes are accompanied by a tsunami. It is presumed that vertical motion is also a necessary prerequisite. Studies of earthquake motion on land, and careful triangulation of seismic waves from similar submarine earthquakes, have led to the conclusion that the generating source, as defined by the initial and subsequent aftershocks, is usually elliptical The net effect of the disturbance is a rapid but discontinuous series of complex motions in several directions Any such (vertical) motion results in a local elevation or depression of the sea surface. This change in surface level disperses rapidly into a train of oscillatory waves, similar in most respects to the waves produced by a stone dropped into a shallow pool *See* EARTHQUAKE, SEISMOLOGY.

Attempts at theoretical analysis of the generation and subsequent history of the waves so formed have met with only qualitative success; most knowledge of their behavior is empirical Model experiments indicate that wave height near the source is related to the amplitude and intensity of the initiating disturbance, and that wave length is related to the dimensions of the dislocation, which may be a hundred miles or more. The initial surface disturbance contains all frequency components down to that associated with the wave length of the disturbance. Those components having wave lengths that are long in comparison to the depth will travel out at nearly the theoretical velocity for infinitely long waves in water of constant depth (about 400–500 mph in the Pacific Ocean), and all shorter components will travel more slowly (Fig. 1). *See* WAVE MOTION IN FLUIDS

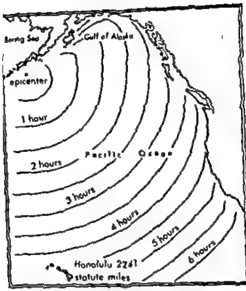


Fig 1 Advance of tidal wave of Monday, Apr 1, 1946, caused by an earthquake with epicenter southeast of Unimak Island (From L. D. Leet and S. Judson, *Physical Geology*, 2d ed., Prentice-Hall, 1958)

The wave train at a later stage in its evolution consists of many individual crests whose wave lengths diminish with time at a fixed range of observation, and with increasing distance from the front of the train at any instant of time (Fig 2). The amplitudes of the individual crests are modulated by a slow beat, which divides the train into groups; the maximum amplitude of each group diminishes slowly with distance from the front and directly as the distance from the source. The highest single crest in the train will be the first near the origin of the disturbance, and its order in the wave train will slowly regress backwards through the first group until at a range of several thousand miles, it may be the sixth or seventh to pass the point of observation. Although the initial height of the disturbance may have been several tens of feet, the highest wave at a radius of several hundred miles will be only of the order of a few feet.

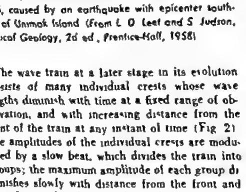


Fig 2. *Idealized* diagram of wave train development from an impulsive disturbance.

On a tsunami's approaching a large island or continental shoreline, however, the combined effects of refraction, resonance, funneling, and interference can cause great local enhancement of wave height, such that watermarks have been observed as much as 40–50 ft above sea level in the wake of great tsunamis. See SHORE PROCESSES. [W.G.V.B.]

Tube, electron

A device in which the conduction of electricity is provided by electrons moving through a vacuum or gaseous medium within a gas-tight envelope. Electron tubes are classified as either vacuum tubes or gas tubes. In a vacuum tube, essentially all gas has been pumped out, while in a gas tube, a small amount of a particular gas is placed in the tube after evacuation and before sealing to alter the characteristics of the tube in a desired manner. In Great Britain, a tube is called a valve.

A tube may provide rectification, amplification, modulation, demodulation, limiting, and a host of other functions when appropriately connected into an electronic circuit. Vacuum tubes are classified as diodes, triodes, tetrodes, pentodes, hexodes, and heptodes, depending on whether they have 2, 3, 4, 5, 6, or 7 electrodes. All tubes must have a cathode to serve as the source of electrons and an anode to collect the electrons. Electrodes in between these two, which serve primarily to control the flow of electrons, are called grids.

In a cathode-ray tube, the electron beam is allowed to pass through a hole in the anode and travel to a fluorescent screen on the inside of the tube face, where it produces a visible pattern. See ELECTRON TUBE; GAS TUBE; VACUUM TUBE

[J.M.R.]

Tube tester

With the large number of tubes available, it is necessary to have available a tube tester that can be used to check to see whether tubes are operating properly. This requires a rather complex instrument, because almost every tube employs a different set of operating voltages, has different base connections, and may fit into one of a number of sockets. A number of tube testers are commercially available ranging from simple devices, which may be used by radio and television servicemen, to more complex devices useful for research laboratories.

Commercial testers. The simplest tube testers are instruments containing a number of different sockets, into which tubes with different bases can be plugged, and with associated switches and controls so that the proper voltages can be applied to the proper terminals. Such tube testers employ a current meter connected to a bridge circuit to indicate the relative mutual conductance of the tube (see VACUUM TUBE). This is done because the mutual conductance is generally the most important operating characteristic of the tube, and if its value is adequate the tube will give proper operation. Usually th[illegible] ter will have bands marked

upon it indicating whether the mutual conductance can be classified as good, fair, or poor. There will also usually be means of checking whether the current emission is adequate and whether there is a short circuit between the filament and the cathode.

More complex tube testers are now available. In one instrument various perforated cards can be inserted, and connections for a tube are made through the perforations. A separate card is provided for each tube type This tester makes it possible to connect each tube in the proper fashion without the possibility of human error.

For precision work more elaborate circuits are available, such as the vacuum-tube bridge This instrument has a number of tube sockets to which connections are made via attached leads so that various tubes can be tested The instrument contains several bridges, which make it possible to measure the resistance, voltage ratio, and transconductance parameters referred to any pair of electrodes The instrument is designed so that transistors as well as tubes used in unconventional and special circuits may be tested, with any one electrode being used as the control electrode. The circuits have large enough current-carrying capacity and sufficient insulation so that transmitting tubes can be tested in addition to receiving tubes and transistors.

Principle of operation. The basic principles involved in the construction of tube testers are quite simple, though the circuits may become complex Each tester will contain one or more power supplies to furnish the proper electrode potentials and transformers for application of filament voltage. For measurement of mutual conductance, a small alternating voltage is applied to the control electrode and the resultant alternating current is read in the appropriate output circuit For measurement of dynamic resistances, an ac bridge circuit is used. Such a bridge circuit reads resistance in the usual fashion For measurement of voltage [illegible] small alter- [illegible] output [illegible] oltages are of opposite phase and are adjusted until the resulting alternating current in the output circuit is zero. The value of the amplification factor is then the ratio of the two voltages

Interpretation of results. The interpretation of results will depend somewhat upon the quantity measured and the particular tube tester used With reference to mutual conductance, it must be remembered that amplification of a tube will in general be proportional to this factor, so that if mutual conductance is low, amplification will be poor It is sometimes possible to observe how rapidly the mutual conductance rises to its final value when the tube is suddenly turned on after having been cold. If the mutual conductance rises rapidly to its final value, with allowance being made for the time required for the filament and cathode to heat up, then the tube is in good shape. If, on the other hand, this rise is slow or erratic, the tube is probably not useful.

A low transconductance reading does not necessarily mean that the tube is inoperative, but that its parameters are deviating more than is considered normal from the average or design center values However, many circuits are designed more or less critically so that even a small deviation of several per cent of this parameter will render the circuit inoperative. Thus the tube tester can only answer questions about how the tube compares with the average. From this point of view the indications of good, fair, or poor commonly used do not always tell whether or not the circuit will function. *See* ELECTRON TUBE. [KRS]

Tuberculariaceae

A family of fungi of the order Moniliales. The conidiophores are short, forming cushion-shaped aggregates, so-called sporodochia, which are like acervuli but superficial. The sporodochia often are waxy or gelatinous There are about 150 genera and 500 species recognized The genera are usually arranged into spore groups, depending on characteristics of the spore, such as the number of cells in the spore, and the shape and pigmentation of the spore *See* FUNGI IMPERFECTI; MONILIALES.

The Hyalosporae have spores which are 1-celled and bright. *Tubercularia* is a genus with large tubercular sporodochia breaking out through the bark of trees *T vulgaris*, with pinkish sporodochia, is the imperfect stage of *Nectria cinnabarina* (Ascomycetes), the coral spot fungus of woody plants

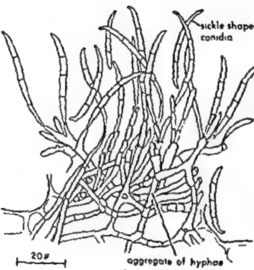
The Hyalophragmiae have spores which are bright and have two or more cells with cross walls (septate) *Fusarium* is a genus with macroconidia which are sickle-shaped slimy spores with pointed ends and cross walls

Many species produce small, usually nonseptate microconidia and some species produce chlamydospores, thick-walled terminal asexual spores The perfect stages belong to such genera of Hypocreales as *Nectria* and *Hypomyces*. About 65 species are recognized Important pathogens are *F. conglutinans*, causing cabbage yellows, *F oxysporum* var. *cubense*, causing Panama disease of banana, *F. vasinfectum*, causing cotton wilt, and *F lini*, causing flax wilt.

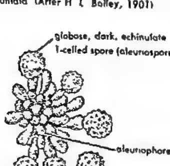
Cylindrocarpon is a genus with macroconidia like those of *Fusarium*, but more nearly cylindrical and with rounded ends. There are 10 species *C. mali* is the imperfect stage of *Nectria galligena* which causes apple and pear canker

The Phaeophragmiae have dark spores containing two or more cells *Exosporium* is a genus with sporodochia which are cushionlike in form (pulvinate) and have spores with several cross septa The species are saprophytes or weak parasites. *E. tiliae* is found on twigs of linden and *E palmivorum* causes leaf-spot of palms.

The Phaeodictyae are a spore group with dark several-celled spores (muriform) *Epicoccum* is a genus with small sporodochia [illegible] spores are



Fusarium lini Sporodochium with sickle-shaped, multicelled conidia (After H L Bolley, 1901)



Epicoccum nigrum Cluster of aleuriophores with aleuriospores (After Saccardo, 1877–1886)

aleuriospores and are flask-shaped, dark, spiny (echinulate), or pitted (foveate) They are originally long and 1-celled but later become muriform [N F B]

Tuberculosis

A chronic infectious disease in man and animals caused by the tubercle bacillus (*Mycobacterium tuberculosis*) It produces many clinical symptoms and a wide variety of lesions (granulomas) with certain histologic characteristics in common—collections of epithelioid and giant cells with a type of necrosis called caseation, in which tissue dies and is formed into a cheeselike substance Human and bovine types of tubercle bacilli are the principal agents of tuberculosis in man Man-to-man transmission is primarily by inhalation of droplet nuclei of sputum Early multiplication of the parasites occurs in alveoli of the lung (Fig 1) and in draining parabronchial and mediastinal lymph nodes Drinking bovine tubercle bacilli in unpasteurized milk (now uncommon in Western civilization) produces primary infection in the gastrointestinal tract and tends to result in extrapulmonary patterns of disease, for example, mesenteric adenitis, cervical adenitis, Pott's disease (tuberculosis of vertebral bodies), and joint disease The avian type of tubercle bacilli causes endemic disease of fowl and swine and is a rare cause of tuberculosis in man. Unclassified atypical tubercle bacilli, commonly photochromogenic with yellow pigment, cause progressive pulmonary disease in man.

Tubercle bacilli. Tubercle bacilli are strongly acid-fast rods (see ACID-FAST STAIN), 1–4 µ in length and 0 3-0.6 µ in diameter, straight or somewhat bent, with parallel sides and rounded ends. These organisms do not form spores and are strictly aerobic (require oxygen for their metabolism). The microorganisms can use a multiplicity of simple carbon compounds as source of energy, and ammonia or amino acids as source of nitrogen for growth They need carbon dioxide for initiation of growth, and some strains require concentrations higher than those available in ordinary air. The optimum conditions for growth are pH 6 5–6 8 and temperature 35–40°C. the minimum generation time is 12–20 hours. A definitive capsule is not formed The hydrophobic and lipophilic nature of the bacterial surface is due to a thin layer of lipids which include high-molecular-weight, acidic polysaccharide esters of mycolic acids These esters of mycolic acid fix neutral red, and account for the serpentine cord formation of virulent strains. Tubercle bacilli are as susceptible to desiccation, heat radiation, and physical agents as other nonsporulating bacteria, but are less injured by mineral acids, alkali, and quaternary ammonium compounds. The principal types of tubercle bacilli are human, bovine, avian, and atypical These are not serologically distinguishable in a qualitative sense since protein, polysaccharide, and lipid antigens give broad cross reactions

Pathogenesis. Acute tuberculosis infection in man and animals produces an exudative, entirely nonspecific type of lesion Development of host

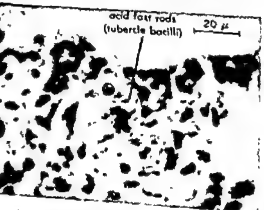


Fig 1. Photomicrograph of a *primary* tuberculous lesion in the lung of a guinea pig, 10 days after air-borne infection. Note cellular accumulation in this enlarged alveolar septum

resistance in chronic forms of the disease produces a proliferative type of lesion, resembling both granulation tissue and a tumor, hence called infectious granuloma. The bacilli multiply intracellularly in nonulcerated lesions and extracellularly in ulcerated lesions such as pulmonary cavities. The bacilli spread locally by direct contiguity, and throughout the body by lymphatic routes, by the blood stream (hematogenous dissemination, which occurs most frequently during the progress of first-infection tuberculosis and results in miliary disease in organs such as the lungs, spleen, liver, kidneys, and meninges), and by tubular means (the contents of a pulmonary ulcer or cavity reach the bronchi and are aspirated into the parenchyma of the lung, initiating new pulmonary lesions or tuberculous laryngitis, tonsillitis, or enteritis; tuberculosis of the kidneys may similarly lead to tuberculous cystitis).

The first infection or primary type of tuberculosis is an acute process, healing or progressing in a relatively short time; the postprimary type (reinfection, either from within or from without) is more chronic because it is associated with a significant, but inadequate, degree of immunity. The progressive primary type of disease is most commonly seen in infants and children, but it may also be seen in adults who have escaped earlier infection. The primary infection may completely heal, but without complete eradication of the parasites; the appearance of postprimary chronic tuberculosis (phthisis), a result of resumed multiplication of the few bacilli which have survived in the primary lesions, may occur several years later. The vast number of cases of tuberculosis constitutes a potential reservoir of infection because the disease may reactivate when the individual is under psychological or other stress (Fig. 2)

Within two or three weeks after primary infection with tubercle bacilli, man and several species of animals develop a delayed tuberculin allergy, an altered tissue reactivity to the bacilli, especially to the bacillary proteins This allergy is not serum-transferable to a normal host, but is transferable by cells of the lymphoid series such as lymphocytes (white blood cells) *See* HYPERSENSITIVITY

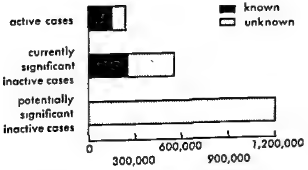


Fig 2 Estimated prevalence of tuberculosis cases in the United States, 1956 (*Division of Special Health Services, U S Dept. of Health, Education, and Welfare*)

This type of hypersensitivity can be established in certain experimental animals against a wide variety of foreign proteins by injection of these proteins in water-in-oil emulsions with whole, killed tubercle bacilli or certain lipopolysaccharide fractions of them. Hypersensitivity to tuberculin indicates previous infection which may or may not have progressed to an extent permitting its detection by clinical or roentgenographic (x-ray) means A negative test almost always rules out infection The blood sera of tuberculous animals contain antibodies against protein, carbohydrate, and lipid fractions of tubercle bacilli and promote phagocytosis by leukocytes. However, such sera have no specific bactericidal or lytic effects on tubercle bacilli, and attempts at passive serum transfer of acquired immunity to tuberculosis have been unsuccessful.

Immunology. The mechanism of specific acquired immunity to tuberculosis is unknown. Experiments in tissue culture have shown that the multiplication of tubercle bacilli is inhibited in mononuclear phagocytes from immune animals Also, much experimental evidence suggests that specific acquired immunity is not necessarily dependent upon tissue hypersensitivity to tuberculin, although the focal necrosis resulting from the hypersensitiveness probably creates an environment which is physicochemically unfavorable for multiplication of the bacilli, particularly since tubercle bacilli are strict aerobes. Partial immunization against tuberculosis has been achieved in experimental animals and in man by the injection of tubercle bacilli killed by various physical or chemical agents or of living strains of attenuated bacilli such as BCG (bacillus of Calmette and Guérin). Some increase in resistance to infection can be conferred with little or no tuberculin allergy by a methanol-soluble fraction of tubercle bacilli

Epidemiology. Tuberculosis has been known in urban civilization since the beginning of recorded history In Western civilization the death rate has been declining since 1870 In America and in most of Western Europe tuberculosis is still the most frequent cause of death between the ages of 15 and 45 years, and for the world at large it is probably the principal cause

The natural progress and the extent of infection in any exposed individual are influenced by many factors, genotypic and phenotypic, such as race, sex, nutrition, and endocrine balance Inhalation of dust containing silica (silicon dioxide) promotes the disease by its local toxic action on the tissue of the host

In the prevention and control of tuberculosis segregation of individuals excreting tubercle bacilli in the sputum, vaccination with BCG, and chemotherapy of primary infections and postprimary disease play important roles

As a result of chemotherapy with isoniazid streptomycin, and *para*-aminosalicylic acid, the mortality (death rate) is rapidly declining in the western world; morbidity (ratio umber of

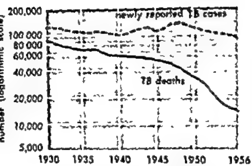


Fig 3 Newly reported tuberculosis cases and tuberculosis deaths in the United States, 1930–1956 (Division of Special Health Services, U.S. Dept. of Health, Education, and Welfare)

... more slowly. The ... there has not ... number of newly reported tuberculosis cases (Fig 3)

Laboratory procedures. Laboratory diagnostic procedures aim to demonstrate tubercle bacilli in pathogenic materials such as sputum, urine, or spinal fluid, by specific staining of smeared material, by artificial cultivation, or by animal inoculation. Artificial cultivation of the tubercle bacilli from pathologic material may be performed on complex media containing egg yolk, potato starch, and glycerin (for example, Jensen's modification of Lowenstein's *medium*), or on better-defined transparent agar media containing long-chain fatty acids, glutamic acid, *mineral* salts, and serum albumin (as a protective agent binding excess fatty acids and other contaminating toxic substances). Guinea pigs, mice, and chick embryos (yolk sac) are employed for diagnostic animal inoculation studies, but cultural methods are usually preferred. *See* BACTERIOLOGY, MEDICAL; EPIDEMIOLOGY; MYCOBACTERIACEAE [G H M]

Bibliography. R. J. Dubos (ed.), *Bacterial and Mycotic Infections of Man*, 2d ed., 1952.

Tube-still heater

The only direct fired equipment used in processing petroleum today. The basic feature of these heaters is that the fluid (liquid, vapor, or a mixture of both phases) is heated in carbon- or alloy-steel tubes which in turn receive their heat from hot products of combustion. The fuel may be liquid, such as heavy fuel oil, or gas, such as natural gas or cracked refinery gas. Burners of the steam, gas, or mechanically atomizing type may be used for liquid fuels, and burners of the simpler ring-type, spider, or center nozzle type for gas fuels. Combination burners designed to handle both fuels are in wide use.

Use of "still" in the name comes from the fact that the first continuous heaters were those for heating the topping still which separated the various fractions of the virgin crude by distillation at atmospheric pressure. Later, to permit distillation of higher boiling fractions, the distillation was carried out under subatmospheric pressure, whence the name vacuum pipe still. With the advent of thermal cracking, the tubular heater served the dual purpose of heating the oil to be cracked to the reaction temperature and of maintaining it under the proper temperature, pressure, and time conditions while providing the required endothermic heat of reaction. Today it is the general purpose tool for supplying heat at the higher temperature levels, that is, over 650–700°F, wherever required in the refinery for distillation, for preheating, for ... d stream, ... case, the tubes are connected in series ... suitable return bends, which may be provided with clean-out openings. Two or more parallel fluid streams may be heated within the same box in order to keep the friction drop through the heater down to desirable limits. Sometimes streams for different services are heated in the same furnace. However, flexibility today usually dictates that the streams be fired and controlled separately. In certain special cases, for example, in steam methane reformers, where large volumes of gas are passed through catalyst filled tubes and pressure drop must be minimized while temperatures must be carefully controlled, all furnace tubes are manifolded in parallel, with all tubes given essentially the same treatment.

Most modern heaters are of one of two types, the box type, usually with horizontal tubes (Fig. 1),

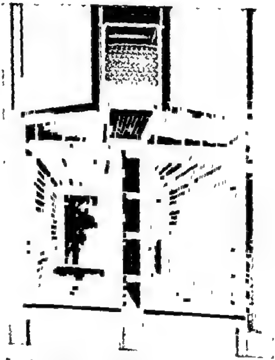


Fig. 1 Double radiant box furnace (The Lummus Company)

or the cylindrical type with vertical tubes arranged in a circle (Fig. 2). Within the radiant section, the combustion zone in which the burners are located, heat is given up to the tubes from four sources: radiation from the flame, radiation from the hot molecules (carbon dioxide and water vapor), radiation from the hot, exposed refractory, and convection from the gases within the chamber. The convection section is basically a recuperative zone in which the flue gases can be reduced in temperature while heating the coldest fluid to obtain economic heat recovery. Higher over-all efficiency may be obtained with air preheaters. In this latter zone, heat is transferred to the air used in combustion by radiation from the hot gases passing through the staggered bank of tubes, by convection from such gases, and by radiation from the hot refractory confining the section. The first two rows to receive the

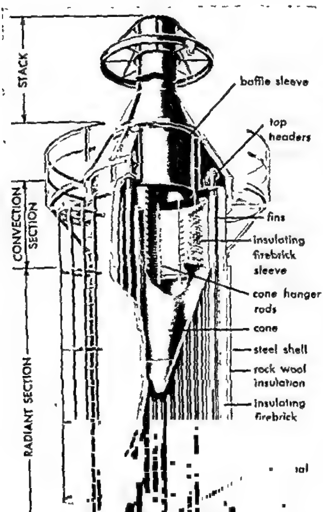


Fig 2. Cylindrical furnace. (Petro-Chem Development

hot flue gases from the radiant section often are called the shield or shock tubes, because, as a result of their position, they also receive heat either directly from the combustion box, as do the radiant tubes, or indirectly from large areas of hot radiating refractory.

The design of tube-still heaters is based as much on experience as on engineering science. The designer must consider not only structural arrangement and heat-transfer rates, but also the effects of fly ash on refractories and metals, strains from heating and cooling, corrosion and erosion, inlet and outlet pressures, methods of operating, and methods of control

The calculation of pressure drop on the fluid side of the tubes is often complicated because the conditions are nonisothermal, and the relative volumes and compositions of the liquid and gas phases may be changing rapidly.

Tube stills have been designed for operating temperatures up to 1900°F, for pressures as high as 4500 psi, and as low as 60 mm Hg. Radiant section heat-transfer flux may be as high as 30,000–40,000 Btu/(hr)(ft^2), though the more common figure is 12,000–14,000 for the cleaner, lighter oils and 8000–10,000 Btu/(hr)(ft^2) for the heavier, dirtier ones. The severe conditions often encountered in furnace tubes were a prime incentive in the earlier development of strong, high-temperature alloys. See FURNACE CONSTRUCTION; HEAT TRANSFER; PETROLEUM PROCESSING. [W.E.L.]

Bibliography: W. E. Lobo and J. E. Evans, Heat transfer in the radiant section of petroleum heaters, *Chem. Eng. Progr.*, 35:743, 1939; W. H. McAdams, *Heat Transmission*, 3d ed., 1954.

Tubiflorales

A large order of the plant subclass Dicotyledoneae including 19 families with 1252 genera and over 19,000 species. This group reaches the culmination of hypogyny (petals and stamens on the receptacle beneath the ovary and free from it) in the Sympetalae and has conspicuous, tubular corollas. The flowers range from regularity in the morning glories (Convolvulaceae) and polemoniums (Polemoniaceae) to irregularity in the mints (Labiatae) and figworts (Scrophulariaceae). Here the irregularity is of the bilabiate type, meaning that the corolla has a two-lipped mouth. This order contains many useful and ornamental plants: sweet potato, Irish potato, phlox, sweet william, heliotrope, forget-me-not, bluebells, verbena, lantana, chaste tree, teak, mints, sage, deadly nightshade, ground cherry, peppers and pimento (*Capsicum*), paprika, egg plant, tomato, tobacco, mullein, digitalis, catalpa, trumpet creeper, sesame, unicorn plant, African violet, gloxinia, thunbergia, and strobilanthes. Dodder (*Cuscuta*) and members of the broomrape family (Orobanchaceae) are parasites, and the butterworts (*Pinguicula*) and bladderworts (*Utricularia*) are insectivorous plants. See ATROPINE; BELLADONNA; DIGITALIS; EGGPLANT; PAPRIKA; PEPPER; PEPPERMINT; [illegible]TO; PO-

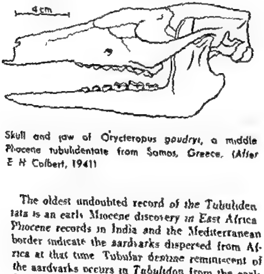
TATO, IRISH; SAGE; SPEARMINT, TOBACCO; TOMATO; see also DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM [P.D.S.]

Tubulidentata

An order of mammals which contains a single species, the aardvark (*Orycteropus afer*) of Africa. Fossil aardvark remains have been found in Europe and Asia, but all belong to the living genus. The aardvark is about the size of a pig. It feeds on termites, and its structure is specialized accordingly, with degenerate jaws, greatly reduced and modified teeth, and powerful forelimbs for digging. The aardvark resembles the anteaters in many ways, and was long regarded as an edentate. More detailed recent study has shown that these resemblances are the result of convergence caused by similar food habits. The aardvark is apparently an isolated and highly specialized survivor of the primitive protoungulate stock, a group of mammals that was prominent and widely distributed during the Paleocene and Eocene. *See* EUTHERIA, MAMMALIA; *see also* EDENTATA. [D.D.D.]

Tubulidentata fossils

Little is known of the origin and history of this peculiar group of ant-eating mammals, represented today solely by the aardvark of Africa. These powerful diggers are characterized by strong jaws bearing peglike, cement-covered, enamelless teeth made up of tubular dentine. Despite specializations for a fossorial mode of life, the skeleton as a whole retains a striking similarity to that of the early condylarths from which the aardvarks may have been derived. *See* CONDYLARTHRA.

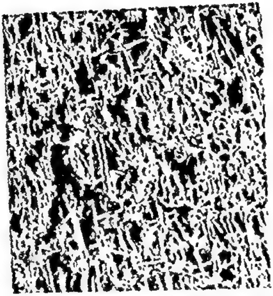


Skull and jaw of *Orycteropus gaudryi*, a middle Pliocene tubulidentate from Samos, Greece. (*After E. H. Colbert, 1941*)

The oldest undoubted record of the Tubulidentata is an early Miocene discovery in East Africa. Pliocene records in India and the Mediterranean border indicate the aardvarks dispersed from Africa at that time. Tubular dentine reminiscent of the aardvarks occurs in *Tabulodon* from the early Eocene of North America, but the fragmentary material does not permit further comparison. [R.H.T.]

Tufa

A spongy, porous limestone formed by precipita-... pring and river waters, also



Calcareous tufa deposited on plant stems. Note high porosity and spongy character. Recent. (*From F. J. Pettijohn, Sedimentary Rocks, 2d ed., Harper, 1957*)

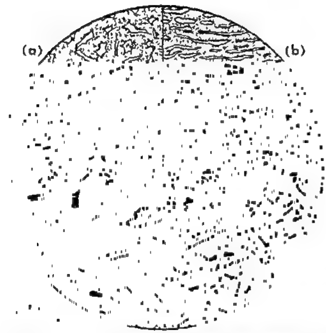
known as calcareous sinter. Calcium carbonate commonly precipitates from supersaturated waters on the leaves and stems of plants growing around the springs and pools and preserves some of their plant structures. Tufa tends to be fragile and friable. Tufa deposits are limited in extent and are found mainly in the youngest rocks, Pleistocene or Recent. See LIMESTONE, THERMAL SPRING, TRAVERTINE. [R.S.]

Tuff

Consolidated volcanic ash, composed largely of fragments (less than 4 mm) produced directly by volcanic eruption. Much of the fragmental material represents finely comminuted crystals and rocks. Much, however, was ejected as blobs of liquid lava which rapidly cooled to form volcanic glass. Lava brought from high pressure within the earth, to low pressure at the volcanic vent, may be explosively disrupted into small masses and droplets, because of sudden expansion of abundant dissolved gases. Expansion and solidification to glass may be more or less contemporaneous so that highly vesicular bodies are rapidly shattered to produce vast quantities of tiny particles (ash). Upon consolidation volcanic ash yields tuff. With increase in size of fragments, tuff passes into lapilli tuff, volcanic breccia, and agglomerate. *See* PYROCLASTIC ROCKS, VOLCANIC GLASS.

In general the quantity of dissolved gas in lava determines the explosiveness of the eruption and therefore, the size of fragmental particles. The coarser and heavier fragments settle near the volcanic vent whereas the finer ash may be ... [illegible]

... (shales, sandstone ... limestone ... [illegible]



Volcanic tuff. (a) Rhyolitic vitric tuff with shards of glass in matrix of dustlike particles. (b) Welded tuff showing deformed and flattened glass shards. (c) Crystal tuff. (d) Lithic tuff.

fallen on dry land, is readily washed away by rain and streams and deposited in the sea to form well-stratified rocks. Evidence of such transportation lies in part in the rounded character of the larger fragments. Rocks thus formed are commonly termed hybrid tuffs. *See* SEDIMENTARY ROCKS; *see also* MARINE SEDIMENTS.

Tuff may be referred to as rhyolitic tuff, trachytic tuff, or andesitic tuff, depending on the approximate bulk composition. These types may be classed as vitric, crystal, or lithic according to whether the principal constituent is glass, crystals, or rock fragments, respectively. These types are gradational

Vitric tuffs are products of explosive eruption of liquid lava. Comminution of the highly vesicular glass yields irregular fragments (shards) more or less bounded by concave surfaces which represent the walls of ruptured gas bubbles Numerous vesicles may remain intact in the larger fragments (pumice). Unaltered glass shards may be clear and colorless or clouded with black magnetite dust Basaltic glass, which is relatively uncommon, is yellow to brown *See* OBSIDIAN.

Vitric tuffs are generally more characteristic of highly explosive eruptions, and they commonly occur at greater distances from the volcanic source They are usually rhyolitic, less commonly dacitic, andesitic, and trachytic, and rarely basaltic.

Occasionally highly gas-charged, viscous lavas erupt with such violence that hot, dense, incandescent clouds of finely divided particles are formed. These *nuées ardentes* emerge from the flanks of the volcanic cone or spill over the crater rim and descend the slope as an avalanche. As the interstitial steam and hot gases escape and the vitric material settles and compacts, individual glass shards are flattened and fused together to form a welded tuff or ignimbrite. *See* VOLCANO.

Crystal tuffs are chiefly products of explosive eruption of lava in which abundant crystals have already developed. Though originally well formed, many of these crystals may be broken by ejection Many may retain thin films of glass which represent viscous lava adhered to crystal surfaces during eruption. Highly fluid lava may be wiped completely free of the crystals by atmospheric friction.

Less abundant constituents of crystal tuffs are crystals and crystal fragments torn from solidified rock.

Lithic tuffs are composed chiefly of more or less angular fragments produced by extensive shattering of solid rock during volcanic eruption. The source rocks are chiefly slightly older lavas and volcanic deposits which are disintegrated during subsequent eruptions. Other rock types (plutonic, sedimentary, or metamorphic) may be present in small quantities. These materials are probably torn loose from the walls of the conduit along which the lava breaks through to the surface.

Volcanic tuff and ash are highly susceptible to alteration. The glassy constituents commonly devitrify or crystallize to extremely fine-grained aggregates of silica and feldspar Rhyolitic glass is commonly converted to clay minerals (largely montmorillonite) to form the rock bentonite Brown basaltic glass is converted to yellow palagonite as it combines with water The breakdown of crystal constituents in tuffs is similar to that of the constituents in most igneous rocks. *See* IGNEOUS ROCKS [C.A.C.]

Tularemia

A widespread infection of wild rodents (48 species), endemic in the Americas, Europe, and Asia. It is maintained through insect reservoirs and vectors. The causative organism, isolated from a plaguelike disease among ground squirrels in Tulare County, California, was named *Pasteurella tularensis* (*Bacterium tularense*). It varies in size and shape, is filterable, and is delicate in structure The organism is not motile and has no flagella or capsules It grows luxuriantly on hemoglobin-cystine agar or thioglycolate–heart infusion agar but will not grow on ordinary media. It finds the best-balanced environment within the cells of the host, either mammals or the chorioallantoic membrane of embryonated hen's egg. Though it can survive in culture for years, it is destroyed in 10 min at 55–60°C The antibiotics streptomycin and tetracycline also destroy the organism and have been effectively used in treatment of the disease Early treatment with these antibiotics has reduced the case fatality rate

The infection may be generalized or it may be localized in the eye, skin, lymph nodes or respiratory or gastrointestinal tracts Recovery is followed by a substantial immunity Between 1914 and 1955 the U.S. Public Health Service received reports of 10,865 cases, the sources were animals

such as rabbits, terrestrial rodents, game birds, sheep, deerflies *Chrysops discalis* (deerfly fever), and ticks *Dermacentor andersoni* (glandular tick fever) The disease may also be contracted by ingestion of water from streams contaminated with the organism It is not contracted by the ingestion of adequately cooked game. The disease has also occurred in laboratory personnel working with cultures of *Pasteurella tularense*. Ticks in all stages from larvae to adults are transmitters; the organism can be transmitted transovarially to the offspring.

Since the reservoirs cannot be destroyed, and the incidence of the disease in the United States has not warranted consideration of any vaccination program the only known preventive measure is education of sportsmen, butchers, and those who live in endemic regions It is advisable not to skin diseased animals Persons skinning wild animals, especially rabbits, should wear rubber gloves In the U S exposure to infected wild animals is erratic, and it is not regarded as a special occupational hazard to animal trappers. Certain laboratories interested in tularemia have worked on immunization from time to time, but not with the idea of any large-scale vaccination There is no commercial vaccine production, nor any licensed vaccine producer, no present source to supply any other than a very special demand. In Russia the disease is more prevalent, and it constitutes an occupational hazard to certain groups such as the fur trade A vaccine is made there with living avirulent strains and there are vaccination programs. For taxonomy see BRUCELLACEAE [K.F.M.E.]

Bibliography R J Dubos (ed), *Bacterial and Mycotic Infections of Man*, 3d ed , 1958

Tulip tree

A tree, *Liriodendron tulipifera*, also known in forestry as yellow-poplar, belonging to the magnol-

siar

ma

ft

whi

min ...ned and covered by two valvate scales stipular scars encircling the twig, an aromatic odor resembling that of magnolia, chambered white pith, and cone-shaped fruit which is persistent in winter See BUD (BOTANY); FRUIT (BOTANY), LEAF (BOTANY), PITH. The name tulip refers to the large greenish-yellow and orange-colored flowers See FLOWER (BOTANY) The wood is light yellow to brown, soft, and easily worked, hence the common name yellow-poplar, a misnomer It is used for construction, interior finish, boxes, crates, baskets woodenware, excelsior, veneer wood and sometimes for paper pulp Because of its wide natural dimensions, the tulip tree often yields lumber as much as 60 in wide which is val-

A leaf, a bud, and a twig of tulip tree, *Liriodendron tulipifera* The bud shows the two valvate scales below which are a leaf scar and a stipular scar (From A. H Graves, *Illustrated Guide to Trees and Shrubs*, rev. ed , Harper, 1956)

uable for backs of wardrobes and other articles of furniture. The standing saw timber has been estimated at 13,600,000,000 board ft, three-quarters of which is below the Mason and Dixon line in the United States. The annual cut is about 700,000,000 board ft. See TREE. [A.H.G.]

Tumbling mill

A grinding and pulverizing machine consisting of a shell or drum rotating on a horizontal axis. The material to be reduced in size is fed into one end of the mill. The mill is also charged with grinding material such as iron balls As the mill rotates, the material and grinding balls tumble against each other, the material being broken chiefly by attrition Tumbling mills are variously classified as pebble, ball, or rod depending on the grinding material, and as cylindrical, conical, or tube depending on the shell shape. See CRUSHING AND PULVERIZING; GRINDING MILL; PEBBLE MILL [R.M.H.]

Tumor

A term ordinarily used to denote a new growth of benign or malignant nature. It is applicable to any swelling or enlargement Such neoplasms occur in many plant and animal species, and the vertebrates, including man, present a wide range of varieties.

Tumor cells, although they may retain some of the charac...

m

p

... tumors have several distinguishing features They tend to be undifferentiated, or anaplastic, so that they resemble embryonic cells more than their fully differentiated parent forms. In many instances, anaplasia is so marked that the origin of the cells cannot be established with certainty; anaplasia and malignancy usually have a

positive correlation. Malignant cells show higher growth rates than do normal cells and much of this growth may be abnormal, so that bizarre shapes, sizes, mitoses, and nuclei are seen. This pleomorphism also correlates with degree of malignancy, in most cases. *See* ONCOLOGY.

Malignant cells have the power to spread, or metastasize, to other tissues and organs. Ultimately this feature, plus the others, will cause death unless treatment is successful.

The many types of malignant tumors show great variations in the above features, and they also have individual characteristics as well.

Benign tumors tend to be differentiated, have slower growth rates, have less abnormality in cells, and are usually encapsulated. In addition, they do not metastasize. Death does not result, except in the infrequent case when physical pressure damages vital organs.

In most cases the exact causes of neoplasia are unknown, despite the most intensive investigative efforts. However, certain relationships to predisposing factors are well documented These include age, sex, race, geographic location, occupation, diet, and pre-existing diseases

The nomenclature of tumors is based on histologic classification. Carcinoma refers to any malignancy derived from epithelial tissue, sarcoma is used for those growths arising from muscle, bone, connective tissue, and other mesodermal structures. In most cases when the parent tissue is known, a compound name is used, as in lipoma, to indicate a benign tumor of fatty tissue, or liposarcoma for a malignant one. The most common benign tumors are those of the skin, such as nevi, and connective tissues, such as fibromas and lipomas. The malignancies with the highest incidence rates are those of skin, breast, stomach, cervix, prostate, and colon, in addition to the atypical neoplasias of the blood, such as leukemias

[E.C.ST.]

Tumor viruses

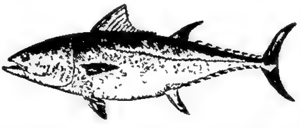
The importance of viruses in the cause and development of malignant tumors is slowly becoming widely recognized. There is a large group of tumors in which a virus has been implicated as the etiologic agent.

The following species are recognized as having virus tumors: (1) frogs, leopard-frog kidney carcinoma; (2) fowl, leukosis complex and some chicken sarcomas; (3) rabbits, rabbit fibroma and papilloma; (4) mice, a number of leukemias and the Bittner milk factor which is capable of producing mammary cancer in mice but is dependent on the hormonal and genetic constitution of the animals, (5) dogs, oral papillomatosis: and (6) man, warts. Many of these diseases are self-limiting and therefore not cancers, but others fulfill all the criteria for true malignancy. Viruses have been considered as one of the possible etiological agents in cancer of man, but since cancers are species specific, direct proof will be very difficult to obtain *See* AVIAN LEUKOSIS; MOLLUSCUM CONTAGIOSUM; MOUSE VIRAL LEUKEMIAS; MYXOMATOSIS, INFECTIOUS (RABBITS); NEOPLASIA; ROUS SARCOMA; VIRUS; WART

[A.E.M.]

Tuna

A name applied to several large fishes of the family Thunnidae, all marine and excellent sport and food fishes. They occur throughout the world in the warm and temperate seas, but are considerably more common in the Pacific than elsewhere When used alone the name tuna generally indicates *Thunnus thynnus*, also called the bluefin and horse mackerel. This tuna attains a length of 14 ft, and a



The tuna, Thunnus thynnus; length to 14 ft (From E L Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

weight of 1600 lb. It travels in large schools and preys on other fish. Albacore, bonito, skipjack, and yellowfin are names applied to others of the tuna group. The average annual world harvest of all tuna is about 675,000,000 lb. American vessels account for about one-half of this catch. *See* PERCIFORMES.

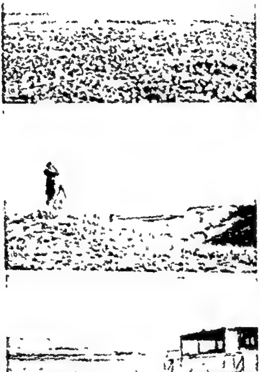
[J.D.B.]

Tundra

A vegetation composed principally of mosses sedges, and lichens, occupying the margins and some of the islands of the Arctic Ocean It grows on cold, waterlogged ground where the subsoil is permanently frozen and the surface is thawed for only a few weeks during the summer Tundra is a transition vegetation between the boreal forest and the barren polar deserts Three general zones are recognized The bush tundra close to the forest margin contains, in addition to the typical association, dwarf trees or bushes growing close to the ground The grass tundra is a more or less continuous mat of sedges and mosses extending northward. Desert tundra occupies small sheltered spots in the polar desert (See illustration on next page.)

Insects are the most numerous animals of the tundras. Swarms of mosquitoes and other biting insects infest the tundra areas in summer The Eurasian reindeer and its wild cousin, the North American caribou, migrate from the forest into the tundra in summer, followed by wolves, foxes, and other predators Small herds of muskoxen inhabit patches of desert tundra in North America

[F.M.D.]



(a) Barren polar desert with scattered small spots of desert tundra in the foothills of northern Alaska (b) Bush tundra, low-growing birch and willow, in foothills of northern Alaska (c) Grass tundra on rolling terrain in Arctic slope of northern Alaska (Alaskan Branch, USGS)

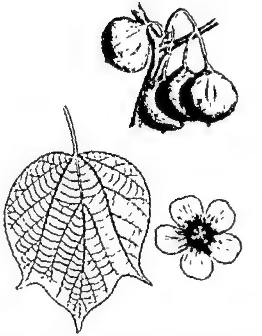
Tuner

A device containing circuits that can be tuned to the carrier frequency of a desired transmitter in the frequency range for which that tuner is designed. The tuner serves to select the desired carrier frequency while rejecting the carrier frequencies of all other stations that may be on the air at that time A tuner for a television receiver ordinarily contains only the radio-frequency amplifier, local oscillator and mixer stages A radio tuner usually contains, in addition, the i f amplifier and second detector stages, so that it can feed an audio signal directly into a separate audio amplifier and loudspeaker system *See* RADIO RECEIVER; TELEVISION RECEIVER

[J.M.R.]

Tung tree

This plant, *Aleurites fordii*, a species of the spurge family (Euphorbiaceae) and native to central and



Leaf, flower, and fruit of tung-oil tree (*Aleurites fordii*) (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

western China, is the source of tung oil The trees have been introduced successfully into the southern part of the United States The globular fruit has 3-7 large, hard, rough-coated seeds containing the oil, which is expressed after the seeds have been roasted Tung oil is used to produce a hard, quick-drying, superior varnish which is less apt to crack than other kinds The foliage, sap, fruit, and commercial tung meal contain a toxic saponin, which causes gastroenteritis in animals that eat this material. *See* DRYING OIL, GERANIALES, VARNISH.

[P.D.S.]

Tungstate

One of a group of compounds containing tungsten in the 6+ oxidation state and derived from tungstic acid, H_2WO_4. Both the normal tungstates such as Na_2WO_4 and the molybdates form condensed ions such as isopolytungstates. These isopolytungstates are usually classified into metatungstates and paratungstates Examples of these compounds have been formulated as $Na_6H_2W_{12}O_{40} \cdot xH_2O$ and $Na_{10}W_{12}O_{41} \cdot xH_2O$, respectively.

Heteropoly compounds in which another anhydride of an element such as silicon combines with the tungsten to form polymers are al- ... dium sil ... of this t ... whereas ...e normal tungstates are sparingly soluble with the exception of the salts of the alkali metals and magnesium

Tungstates are used to flameproof fabrics, and in the manufacture of phosphorescent screens *See* TUNGSTEN

[E.E.WR.]

Tungsten

Element number 74, tungsten, W, is the heaviest element in group VIb of the periodic table. Physically, tungsten is a body-centered cubic metal with atomic weight 183.92. It has a silver to gray metallic luster; high melting point, 3410°C (highest of any metal); low vapor pressure; high hardness; high density, 19.3 g/cm³; and superior high-temperature strength. Chemically, tungsten has oxidation states from 2+ to 6+. The lower states are basic, and the higher are acidic. Natural isotopes and their abundances are 180 (0.2%), 182 (25.8%), 183 (14.2%), 184 (30.6%), and 186 (29.2%).

Uses. Ferrous alloys consume 40% of the tungsten mined. When added to iron or steel, tungsten improves high-temperature strength and hardness. Ferrotungsten, made in electric furnaces using carbon or silicon as reducing agents or by the aluminothermic or silicothermic method, is used for high alloy additions (*see* FERROALLOY). Tungsten concentrates may be charged directly into the steel melt for alloys of lower tungsten content. Nonferrous alloys (using 6% of the tungsten mined) are the cutting-tool, hard-facing type, tungsten-chromium-cobalt alloys. Silver- or copper-tungsten alloys are used for heavy-duty electrical contacts. Alloys for jet engine and missile applications are under development and will be of future importance. High density, tungsten-copper-nickel alloys have been utilized for radiation shielding. Tungsten carbide (38%) has replaced diamond for many die and drilling applications. It is one of the best hard tool materials, retaining its properties at elevated temperatures. Pure tungsten metal as wire, rod, and sheet (15%) is important to the electric lamp, electronics, and electrical industries. Other applications are for welding rod, x-ray targets, lead wires, cathodes for power tubes, and automotive and aircraft distributor points.

Tungsten chemicals (1%) have several minor industrial applications. Tungstic oxide and sodium tungstate are important to the metallurgical processing of pure tungsten, and are available commer-
[illegible]

for fabrics, ink pigments, analytical reagents, and substances used in the glass, ceramic, and tanning industries.

Occurrence. Commercially important tungsten minerals are wolframite, $(FeMn)WO_4$; ferberite, $FeWO_4$; hübnerite, $MnWO_4$; and scheelite, $CaWO_4$. Deposits found in China, Burma, Korea, South America (Bolivia), Europe (Portugal), and the United States (California, North Carolina, Nevada, Colorado, Idaho) contain 0.5–2.0% tungstic oxide.

Extraction and fabrication of metal. Concentration of ores to 60–70% oxide is accomplished by crushing, grinding, magnetic or gravity separation, flotation, or special chemical reactions. Wolframite concentrates are converted to soluble sodium tungstate by fusing with sodium carbonate (900–1040°C) or by aqueous caustic digestion. Iron and manganese oxides are removed as insoluble sludge. Acidifying the soluble tungstate solution precipitates tungstic acid. Scheelite concentrates are treated directly with hydrochloric acid to form calcium chloride and insoluble tungstic acid. After washing, the tungstic acid is dissolved in excess ammonium hydroxide, filtered, and evaporated to form pure ammonium paratungstate crystals. Repetition of this operation improves purity.

Heating the paracrystals to 400–500°C produces pure tungstic oxide; heating in the presence of hydrogen reduces the crystals to the blue $(W_{20}O_{58})$ or brown (WO_2) oxide. Further reduction with hydrogen (800–1000°C) produces pure metal powder. Single-step reduction from ammonium paratungstate to metal powder also is possible. The reduction can be controlled to yield a fine powder (0.5–2.5 microns (μ) average particle diameter) or a powder of any coarseness up to 400-500-μ size. Carbon-reduced metal powder of lower purity also is commercially produced.

It is difficult to melt tungsten directly, although it has been successfully vacuum arc-melted and electron-beam-melted into water-cooled crucibles. Production of pure consolidated tungsten generally is done by powder metallurgical methods. Metal powder is compacted in dies at pressures of 10–20 tons/in.²; the pressed shape is strengthened by presintering at 1200–1300°C in a hydrogen furnace. Final sintering to 90–95% of theoretical density is accomplished by passing an electric current through the bar held in a water-cooled bell jar and heating it to 3000°C. Prolonged heating at lower temperatures will also sinter pressed tungsten powder. *See* POWDER METALLURGY.

Sintered bars are worked into rods by heating in a hydrogen-atmosphere furnace to 1600°C and swaging in air. The swaging temperature is reduced as the rod becomes smaller in diameter. When reduced sufficiently, the rod is hot-drawn into wire using tungsten carbide dies, and ultimately diamond dies, with graphite lubrication. Fine wire has ductility in bend and can be coiled into small-diameter lamp filaments. Pure tungsten becomes brittle when recrystallized. It has goo[illegible] h temperature tensile strength but will [illegible] eep

sag) at the operating temperature of lamp filaments. To overcome this tendency, traces of potassium, silicon, and aluminum are purposely "doped" into the metal powder. During sintering at 3000°C, most of the additives are removed, but the residual impurities impart special properties to the recrystallized wire structure that reduce the sagging characteristics. Up to 2% thoria, ThO_2, is added to tungsten to inhibit the grain growth of recrystallized tungsten wire. Fine-grained, recrystallized, thoriated tungsten wire is brittle at room temperature and will sag at elevated temperatures but has exceptional high temperature shock strength and is used for special lamp and electronic applications.

Tungsten can be soldered or brazed, but welding imposes a problem of embrittlement. Machining the metal is very difficult, grinding or cutting by abrasive wheels is more frequent. Tungsten is cleaned by conventional degreasing, by immersing in molten caustic soda, or electrolytically in an aqueous solution of caustic soda.

Properties of the metal. Chemically, tungsten is relatively inert. It is not readily attacked by the common acids, alkalies, or aqua regia. It reacts with a mixture of concentrated nitric acid and hydrofluoric acid. Molten oxidizing salts such as sodium nitrite attack tungsten rapidly. Tungsten does not oxidize at room tem...

only at high temperatures. Carbon, boron, silicon, and nitrogen also form compounds with tungsten at elevated temperatures, hydrogen does not.

Table 1 lists several of the important physical properties of ...

terial tested.

Principal compounds. Tungsten compounds embrace oxidation states of tungsten from 2+ to 6+, the higher oxidation states being most stable. Tungsten chemistry resembles that of chromium and molybdenum which also occupy the same subgroup in group VIb of the periodic table. Tungsten aqueous chemistry is complicated by its tendency to form complex ions.

Oxides. Tungsten forms four, well-defined, stable oxides. Several others have been reported, but their existence is doubtful. Beta-tungsten (reported as W_3O) has not been conclusively identified as an oxide.

Tungsten trioxide (tungsten oxide), WO_3, has yellow triclinic or pseudo-orthorhombic crystals (recently reported to be monoclinic), and it transforms to the tetragonal form above 740°C. It has specific gravity 7.16 and melting point 1473°C, with tendency to sublime. It is reducible to metal powder by hydrogen at 650°C and above, and by carbon at 1000–1100°C. The oxide is insoluble in water or acids, but forms soluble tungstates in alkali solutions. Its heat of formation is -199 ± 1 kcal/mole.

Other oxides include $W_{20}O_{58}$ ($WO_{2.90}$), dark blue, thin monoclinic needle crystals, heat of formation, -193 ± 1 kcal/mole; $W_{18}O_{49}$ ($WO_{2.72}$), reddish-violet, monoclinic, needle crystals, heat of formation, -183 ± 1 kcal/mole; WO_2, tungsten dioxide, brown, monoclinic crystals, specific gravity 12.11. The pure compound is prepared in vacuum or inert atmosphere by the prolonged heating at 950°C of stoichiometric mixtures of W and WO_3. Partial hydrogen reduction of WO_3 yields WO_2 plus a second phase impurity. Its heat of formation is -137 ± 1 kcal/mole.

Acids. Hydration of tungstic oxide produces ...

... in strong alkaline solutions. It precipitates from warm, strongly acid solution and tends to be colloidal if precipitated from

Table 1. Physical properties of tungsten

Melting point		3410°C
Boiling point		5930°C
Density, metal powder		2.0–4.8 g/cm³
sintered at 1200°C		10.0–12.0 g/cm³
sintered at 3000°C		17.0–18.5 g/cm³
swaged rod		17.0–19.2 g/cm³
drawn wire		19.3 g/cm³
Specific heat (20°C)		0.032 cal/g
Latent heat of fusion		45 cal/g
Latent heat of vaporization		1160 cal/g
Vapor pressure at 1727°C		3×10^{-14} atm
	at 3410°C (mp)	5.4×10^{-6} atm
	at 4227°C	5.6×10^{-3} atm
	at 5727°C	0.9 atm
Thermal expansion, linear		$L = L_0[1 + (4.28t + 0.00058t^2) \times 10^{-6}]$, t = °C
Thermal conductivity	at 0°C	0.40 cal/(cm)(sec)(°C)
	at 1227°C	0.28
	at 1827°C	0.30
	at 2227°C	0.30
Electrical resistivity	at 20°C	5.6 microhm-cm
	at 927°C	30.2
	at 1827°C	59.0
	at 2727°C	90.4

Table 2. Mechanical properties of tungsten

Tensile strength, sintered ingot		18,000 psi
swaged rod		50,000–215,000 psi
drawn wire		250,000–600,000 psi
annealed wire		150,000 psi
High temperature tensile strength, 0.028-in. diameter wire	20°C	430,000 psi
	800°C	200,000 psi
	1500°C	50,000 psi
	2800°C	5,000 psi
Yield strength		Approximately 90% or more of the tensile strength
Ductility		0–4% elongation (brittle when recrystallized)
Modulus of elasticity		12.8×10^6 psi (for sintered rod)
		53×10^6 psi (for well-worked wire)
Modulus of torsion		$12.8–21.3 \times 10^6$ psi
Compressibility coefficient		1.67×10^{-6} per ton/in. (smallest value for all the metals)
Hardness, sintered bar		255 Vickers
swaged bar		400–480 Vickers

weakly acid solution. Boiling with strong acid coagulates the colloid.

Hydrated tungstic acid, $H_2WO_4 \cdot H_2O$, is a white compound precipitated from a cold, acidified solution of H_2WO_4. It is more soluble in water than H_2WO_4, it is converted to yellow tungstic acid on boiling in acid solution, and it tends to be colloidal when washed.

Metatungstic acid, $H_6W_{12}O_{40} \cdot xH_2O$, forms fine yellow crystals that are soluble in water. When heated to 100°C, it is converted to H_2WO_4.

Salts Tungstates have the general formula, $M_2WO_4 \cdot xH_2O$. Alkali metal and magnesium tungstates are water-soluble, others are insoluble. Insoluble tungstates are prepared by fusing the metal oxide with WO_3 or by precipitation from sodium tungstate solution.

Sodium tungstate, Na_2WO_4, forms water-soluble, white, rhombic crystals, with specific gravity 4.18, melting point 692°C. When crystallized from aqueous solution it forms a dihydrate. The anhydrous salt is prepared by direct fusion of WO_3 and NaOH

Ammonium tungstate, $(NH_4)_2WO_4$, cannot be isolated from aqueous solution and it evolves NH_3 on heating It is prepared by adding hydrated tungstic acid to liquid ammonia

Metatungstates, $3M_2O \cdot 12WO_3$ xH_2O, are water-soluble salts prepared by dissolving tungstic acid in metal tungstate solutions

Paratungstates, $5M_2O$ $12WO_3$ xH_2O, the structures of which vary with the solution pH, are prepared by precipitation from slightly acid solution. Sodium paratungstate is made by saturating sodium hydroxide solution with WO_3 Ammonium paratungstate, $5(NH_4)_2O$ $12WO_3 \cdot 11H_2O$, is a fine white crystal when produced by slow evaporation, rapid evaporation yields platelike crystals. Prolonged boiling converts the salt to a decahydrate Crystals are water-soluble but decompose in acid or alkali solution. Heating dehydrates the compound and drives off ammonia. Heating in hydrogen reduces the compound to tungsten metal powder

Heteropoly acids Many heteropoly compounds have been identified, and two of them are well known.

Phosphotungstic acid, H_3PO_4 $12WO_3$ xH_2O forms greenish-yellow crystals prepared by evaporation of a solution containing phosphoric and metatungstic acids.

Silicotungstic acid, Si_2O $12WO_3 \cdot 26H_2O$, forms pale yellow, rhombic crystals soluble in water, alcohol, or ether.

Halides and oxyhalides. Numerous halides have been reported. They are generally unstable in air, have low boiling points, and react with water vapor.

Tungsten hexachloride, WCl_6, forms blue to violet, hexagonal crystals, with melting point 275°C, boiling point 346.7°C. specific gravity 3 52. It is soluble in carbon disulfide, it decomposes in water, and it is reducible by hydrogen to lower chlorides or metal. It is prepared by the reaction of dry Cl_2 with metal powder heated to redness; in the presence of moisture, $WOCl_4$ is formed. Other known chlorides are WCl_2, WCl_4, and WCl_5

Tungsten oxytetrachloride, $WOCl_4$, forms red needles, with melting point 211°C, boiling point 227.5°C, specific gravity 11.9. It is soluble in carbon disulfide, and it decomposes in water. It is prepared by the reaction of Cl_2 with W and WO_2 at 400°C and purified by vacuum distillation at 200°C

Other tungsten compounds are the bromides, WBr_6, WBr_5, and WBr_2; and the iodides, WI_4 and WI_2. Known fluoride compounds are the highly volatile WF_6 and WOF_2.

Carbides. Two carbides are known; each can be prepared by heating mixtures of the elements to 1400–1600°C The compounds are gray powders having hardnesses approaching that of diamond They are insoluble in water, but are attacked by concentrated HNO_3-HF solution.

Tungsten carbide, WC, forms hexagonal crystals, specific gravity 15.6, melting point 2900°C

Ditungsten carbide, W_2C, forms hexagonal (close packed) crystals, with specific gravity 17.2, melting point 2850°C.

Other important compounds. These include the carbonyl, nitride, boride, phosphide, silicide, and sulfide.

Tungsten hexacarbonyl, *$W(CO)_6$, is a white volatile* solid that decomposes into W and CO at 150°C It is insoluble in water, and slightly soluble in organic solvents. Three nitrides have been identified, W_2N_3, WN_2, and W_2N. The known borides are WB_2, WB, and W_2B The known compounds with phosphorus are WP_2, W_2P, WP, and W_3P_4.

Well-defined tungsten silicides are WSi_2, W_2Si_3, and WSi_3 Tungsten disulfide, WS_2, forms soft gray crystals with specific gravity 7.5 Found in tungstenite ores, it oxidizes in air, is reduced by hydrogen at 900°C to metal powder decomposes at 1250°C, is insoluble in water, is attacked by fused alkali or HNO_3-HF mixtures, and is prepared by heating H_2S and WCl_6 or by direct reaction of elements at high temperatures.

Ammonium tetrathiotungstate, $(NH_4)_2WS_4$, forms orange crystals by the reaction of H_2S and ammoniacal tungsten acid solutions *See* METALLURGY; TRANSITION ELEMENTS [C.H.T.]

Bibliography K C. Li and C Y Wang, *Tungsten*, 3d ed. 1955, J. W. Mellor, *Comprehensive Treatise on Inorganic and Theoretical Chemistry*, vol 11, 1931, N V Sidgwick, *The Chemical Elements and Their Compounds*, 2 vols., 1950; C. J. Smithells, *Tungsten*, 1953.

Tunicata

A group of marine animals distantly related to the vertebrates and forming a subdivision (subphylum Tunicata, or Urochordata) of the Chordata. Three classes are usually recognized: Ascidiacea—ascidians, or sea squirts; Thaliacea—salps, doliolids, and pyrosomids, and Larvacea—appendicularians, or Copelata. Approximately 2000 species are known. Some are planktonic but the majority are sessile, and most occur at the depths of than 100

fathoms. Maximum dimensions of these animals seldom exceed a few inches. Definitive characteristics include a restriction of the notochord to the tail and posterior body of the larva. The notochord persists in the adult of the class Larvaces. There is an absence of mesodermal segmentation. An outer covering or tunic secreted over the body usually contains some cellulose.

Tunicates feed on minute plankton and organic detritus which is carried into the pharynx by water currents resulting from ciliary or muscular activity. This material is trapped in mucus secreted by the endostyle, a glandular groove in the pharyngeal floor. The gut is usually differentiated into an esophagus, a stomach bearing one or more glandular structures, and an intestine. The intestine sometimes exhibits specialized regions, and digestion is extracellular. A partly open circulatory system is present. Blood is pumped by a tubular heart which periodically reverses its beat, thereby changing the direction of flow in all channels. In some species, high concentrations of vanadium occur in the blood cells and elsewhere. The brain in the adult is a solid dorsal ganglion associated with a neural gland that is sometimes homologized with part of the vertebrate pituitary gland. All forms are hermaphroditic, though male gonads may develop earlier than the ovaries. The eggs, often used in experimental embryological studies, undergo bilateral, determinate cleavage. A swimming tadpole larva equipped with notochord, dorsal nerve cord and sense organs ... metam... and ne... but nc...

Some ... develop into solitary adults, others undergo more or less alternate periods of sexual and asexual reproduction. They often form aggregates or colonies by budding. Regenerative capacity is usually marked.

Direct economic importance of tunicates is slight. Some forms are a nuisance as fouling organisms on ships and buoys. At least six species are taken for human consumption, primarily in the orient. See ANIMAL SYMMETRY; CHORDATA; CLEAVAGE, EMBRYONIC; REGENERATION (BIOLOGY). [D.P.A.]

Bibliography: N. J. Berrill, *The Tunicata*, 1950; P. Brien and H. Harant, *Embranchement des Tuniciers*, in P. P. Grassé (ed.), *Traité de zoologie*, vol. 11, 1948.

Tuning

The process of adjusting the inductance or capacitance, or both, in a tuned circuit, for example in a radio, television, or radar receiver or transmitter, so as to obtain optimum performance at a selected frequency. The tuning procedure that is carried out during manufacture or servicing of equipment generally involves the adjustment of screwdriver-type controls located inside or at the rear of the equipment. These are adjusted in such a way that the main tuning control on the front of the panel will serve to adjust all tuning circuits in the equipment simultaneously when a change of frequency is desired. See RESONANCE (ALTERNATING-CURRENT CIRCUITS). For an application of tuning, see RADIO RECEIVER. [J.M.R.]

Tuning fork

A steel instrument consisting of two prongs and a handle which, when struck, emits a tone of fixed pitch. Because of their simple mechanical structure, purity of tone, and constant frequency, tuning forks are widely used as standards of frequency in musical acoustics. In its electrically driven form, a tuning fork serves to control electric circuits by producing frequency standards of high accuracy and stability.

A tuning fork is essentially a transverse vibrator. The amplitude of longitudinal vibration at the end of the stem is small compared with the amplitude of the transverse vibrations at the ends of the prongs (see illustration). Thus, when the stem of the fork is pressed on a sounding board or a resonance box, the vibrations persist for a considerable time since the small amplitude vibrations transfer energy to the sounding board at a low rate.

A tuning fork vibrating at its fundamental frequency.

Tuning forks are constructed to cover the entire range of audible frequencies from 20 to 20,000 cycles per sec. The frequency of a fork varies approximately as the inverse square of the length and directly as the thickness of the prongs. The sound output of the fundamental frequency of a fork may be reinforced by attaching the stem to an air column type of resonance box having the same fundamental frequency. See VIBRATION. [L.E.K.]

Tunnel

A general term for subterranean passages. Tunnels are used for aqueducts and sewers; for carrying railroad and vehicular traffic under rivers, through mountains, and below cities; and for specific kinds of underground installations, such as hydroelectric plants.

The ... tio... the ... unnel the ...ing, or arch, as the tunnel is being d... Tunnels driven ... in

Tunnels under rivers in soft silts are usually driven with the aid of a short cylinder called a shield, which is pushed through the silt ahead of the excavation to provide advance support of the arch. After excavation has been completed, the exposed tunnel wall is usually lined with concrete. Vehicular tunnel construction often requires the added installation of tile walls, and lighting and ventilation equipment

Excavation. *This is the most costly and hazardous* operation in tunneling Because the nature of the material through which the tunnel is to be driven determines driving methods, the need for supports, and therefore the costs, it is standard procedure to make a thorough geological survey of the subterranean strata to be encountered This consists of a study of structural characteristics, mineral composition, and hazards. Structural characteristics, such as the frequency and direction of rock joints and planes, will determine the need for and type of arch supports. The mineral composition is significant in forecasting behavior of the tunnel faces after excavation For example, anhydrite when exposed to water during tunneling changes into gypsum with a resulting swelling of the rock and deformation of the tunnel. Hazards, such as faults, indicate the presence of crushed rock fragments or water channels. The former situation requires extensive tunnel supports; the latter, pumping and possible advance grouting For a discussion of grouting, *see* CONSTRUCTION METHODS

The required geological information is obtained from a review of geological literature and maps of the area and from field explorations. Seismic and electrical soundings are taken to establish the depth of rock from the surface Drilled core holes provide samples of the material to be encountered during driving of the tunnel. Topographic surveys made on the ground or from the air are sometimes used to forecast the underground rock contours and character.

Such studies, while *invaluable in selecting a tunnel* route bypassing certain localized hazards can forecast exactly neither all such hazards nor the pressures to be developed on the tunnel face during excavation Exact information can be obtained only by driving small pilot tunnels at the proposed site

Excavation of tunnels through rock proceeds by repeating the following cycle of operations for the length of the tunnel: drilling blast holes, loading the holes with explosives, setting off the explosives, exhausting the blast fumes, and mucking or removing blasted rock.

Drilling. Blast holes are drilled by compressed-air-operated rock drills mounted on a carriage called a jumbo. This carriage may consist of several platforms supported on a tractor, truck, or track-*riding gantry.*

Each drill makes several holes which, together with those made by other drills, form a blast-hole pattern. The blast-hole pattern consists of an inner group, called cut holes, an intermediate group, called relief holes, and finally an outer row called

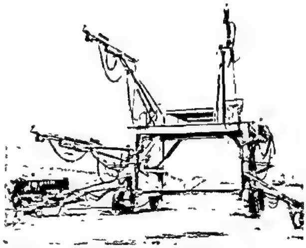


Fig. 1 Drill jumbo. *(Gardner-Denver Co)*

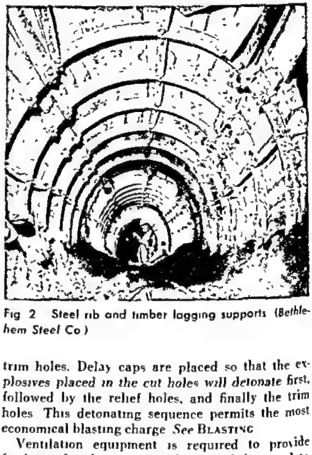


Fig 2 Steel rib and timber lagging supports *(Bethlehem Steel Co)*

trim holes. Delay caps are placed so that the explosives *placed in the cut holes will detonate* first, followed by the relief holes, and finally the trim holes This detonating sequence permits the most economical blasting charge *See* BLASTING

Ventilation equipment is required to provide fresh air for the crews at the tunnel face and to eliminate blast fumes Blowers, set up at the entrance of the tunnel, force fresh air to the tunnel face through large-diameter ducts These blowers are temporarily reversed when used to exhaust the blast fumes

Mucking The rock or earth is removed by special conveyor loaders or power shovels equipped with short booms Rock is hauled from the face by means of special trucks or mine cars

Roof supports Supports must often be provided to support loose rock at the roof Long wedging bolts are used to anchor large blocks of rock to sound rock. Arch supports for scaly or heavily fragmented rock consist of steel or timber ribs with steel or timber members, called lagging, which span the area between the ribs

The design of arch supports requires an estimate of the vertical and lateral pressures expected on the support These will vary with the material encountered Because a natural arch forms above the excavated tunnel, the vertical load will never be as

many times greater In such extreme cases, the tunnel cross section is made circular and circular ribs are used for wall support

To save time, it is preferable to erect arch supports at the same time as drilling blast holes for a new round or advance. However, when the exposed roof cannot support itself until drilling has been

of the face must be excavated at a time to permit more rapid setting of supports.

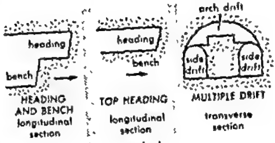


Fig 3. Tunnel-driving methods

Tunnel-driving methods. The amount of face that can be taken out at one time determines which of several tunnel-driving methods to select. These include the full-face, the top-heading, the heading and bench, and the drift methods

With the full face method, the most common and fastest, the tunnel is blasted out full size at each round of blasting The required conditions for full face operations are a rock type and tunnel section that permit self-support until after mucking has been completed In the top-heading method the tunnel portion just below the arch, the heading, is driven full length first, after which the bottom portion, or bench, is removed separately In the heading and bench method, the top portion is also driven first but only a round ahead of the bench This method is often preferred to the top heading method because there is less time for the roof loads to build up The drift method is used where the rock is so poor that the face must be attacked in more than two steps There are two common versions of the drift method the side drift and the multiple drift In the side-drift version two small tunnels are driven at the side of the projected tunnel Steel or timber posts or concrete abutments for the arch supports are then installed, after which the balance of the tunnel face is opened up The multiple drift consists of excavating the arch drift after the side drifts setting arch supports, and then removing the remaining core of rock.

Some tunnels in soft but sound rock, such as limestone or shale, can be excavated by means of special boring machines which bore out a full cross section of the tunnel This eliminates the standard drilling and blasting methods previously described

Driving of tunnels through soft (nonrock) ground, as in vehicular tunnels under rivers, requires a completely different approach The certainty of almost immediate arch collapse requires the installation of continuous circular ring supports which abut each other for the full length of the tunnel The circular supports are installed under the cover of a tunnel shield This is a steel cyl-

a full jack stroke has been made, the jack piston is retracted and new rings are added The material ahead of the shield is either pushed ahead or squeezed into the inside of the tunnel, where it is removed by special digging tools

When the material being driven through is of fluid consistency, the inside of the tunnel must be kept under air pressure to balance the fluid pressure and prevent an uncontrolled flow of material into the tunnel Under such conditions, a bulkhead with a door is placed just inside the shield edge to control the amount of fluid material squeezing into the tunnel. Tunnels driven under air pressure rarely exceed 110 ft in depth, because workmen cannot work under greater pressures Men and materials must enter and leave the tunnel through air locks

The high cost of constructing an underwater tunnel under air pressure has led to the use of the floated tunnel or trench construction method Under this method all work is done from the surface. A trench is first dredged under the river bed, then cylindrical tunnel sections, of either steel or reinforced concrete, are floated over the trench and sunk into place. The sections are joined together underwater by divers, covered over with fill, and then pumped out for completion of work inside the dry sections

Tunnel linings. Concrete lining of a tunnel is required for several reasons In water-supply tunnels, it provides better flow and therefore increases capacity In vehicular tunnels, it is necessary for appearance and safety A lining may also be required to help the arch ribs and roof bolts resist the vertical and lateral pressures that may develop as the rock adjusts itself to new load conditions caused by the driving of the tunnel The surface of the concrete lining is formed by traveling steel forms 20–100 ft in length These forms may be either telescopic or nontelescopic.

In the telescopic system, which is standard for long tunnels, one set of forms is collapsed over a carriage which passes it under another set of forms in position By providing enough sets of telescopic forms it is possible to place concrete continuously

Nontelescopic forms must move as a unit. They are less expensive than telescopic forms, but con-

crete placing must be discontinued until the concrete gains sufficient strength to permit removal and reuse of the forms.

The concrete is placed between rock and form by compressed-air placers or piston-type pumps, which force the concrete through pipes to the top of the arch above the form. From there, the concrete flows down around the form For long tunnels, the concrete is mixed near the form in the tunnel in a traveling mixer. Concrete materials are brought to the mixer by trucks or mine trains. After concreting is done, it is usually necessary to pump cement grout through small pipes left in the lining to seal any voids or gaps between rock and concrete.

[W.H.]

Tupelo

A tree belonging to the genus *Nyssa* of the dogwood family (*see* UMBELLALES). The most common species is *N. sylvatica*, variously called pepperidge, black gum, or sour gum, the authorized name being black tupelo. Tupelo grows in the easternmost third of the United States, southern Ontario, and Mexico. In moist soil this tree usually ranges from 60–80 ft in height and 2–3 ft in diameter, but exceptional trees may be 110 ft tall and 5 ft in diameter. It can be identified by the comparatively small, obovate, shiny leaves, by branches which develop at a wide angle from the axis, and by a chambered pith. *See* LEAF (BOTANY); PITH. The fruit is a small, blue-black drupe, a popular food for birds. *See* FRUIT (BOTANY). The wood is yellow to light brown in color and hard to split because of the twisted grain *See* WOOD (ANATOMY AND IDENTIFICATION); XYLEM. The general increase in value of all woods has resulted in greater use of tupelo for boxes, baskets, and berry crates, and as backing on which veneers of rarer and more expensive woods are glued. It is also used for flooring, rollers in glass factories, hatters' blocks, and gun stocks The combined saw timber stand of black tupelo and water tupelo, *N. aquatica*, has been estimated at about 20,000,000,000 board ft. The average annual cut is about 300,000,000 board ft, over half being black tupelo. *See* TREE

[A.H.G.]

A leaf, fruit cluster, bud, and twig of black tupelo, Nyssa sylvatica (From A H. Graves, Illustrated Guide to Trees and Shrubs, rev. ed, Harper, 1956)

Turbellaria

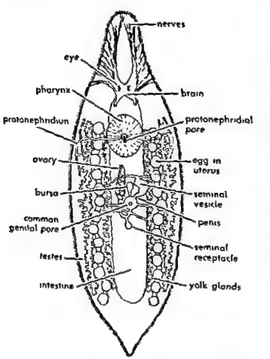
A class of the phylum Platyhelminthes commonly known as the flatworms. These animals are chiefly free-living and have simple life histories. The bodies are elongate and flat to oval or circular in cross section Their length ranges from less than 1 millimeter to several centimeters, but may exceed 50 centimeters in land planaria. Large forms are often brightly colored. Smaller forms may have black, gray, or brown parenchymal pigment or may be white or transparent except for the color of ingested food and symbiotic algae. This class numbering some 3400 described species is ordinarily subdivided into the orders Acoela with 200 species; Rhabdocoela, 1100 species; Alloeocoela, 350 species; Tricladida, 1000 species; and Polycladida, 750 species.

Economics and ecology. Turbellaria are not economically important but have proved valuable in the study of such fundamental biological problems as regeneration, metabolism, axial gradients, evolution, and adaptations to parasitism. Although widely distributed in fresh and salt water and moist soil, they are usually overlooked because of their generally small size, secretive habits, and inconspicuous color. They are seldom eaten by other animals but frequently feed on one another and may harbor commensals and parasites, chiefly Protozoa and Nematoda. Some species are themselves parasites or commensals on other aquatic invertebrates.

Morphology and physiology. Locomotion is by gliding, swimming, or muscular movements of the body wall Respiration takes place by diffusion through the body wall since respiratory organs are lacking The cellular or syncytial epidermis is usually covered with cilia and may contain the nematocysts of ingested coelenterates Gland cells are of frequent occurrence, producing adhesive substances, mucus, and two types of rod-shaped secretions or rhabdoids. These are the commoner rhabdites and the longer and slenderer rhammites. Usually there is a basement membrane beneath the epidermis, inside which lies an outer circular and an inner longitudinal muscle layer, sometimes with a diagonal layer between The muscular system also includes both parenchymal and organ muscles Between the body wall and internal organs lies the parenchyma, a more or less compact network of mesenchyme cells Plasma in its interstices may function as a circulatory fluid.

Digestive system The mouth, on the midventral

pouch, has a limiting membrane except at its base, and is highly protrusible. The bulbous pharynx lies in a shallow pouch and is surrounded by a limiting membrane. It is eversible and occurs in two chief forms, the cask-shaped or doliiform located at the anterior end of the intestine parallel to the main body axis and the globular or rosulate lying ventral to the intestine and perpendicular to the main body axis. A short esophagus is often present and polypharyngy may occur. The intestine may be unbranched or a two, three, or many branched sac with or without diverticula. It is lined with tall epithelial cells and generally has no anus. Turbellaria are carnivorous, digestion is largely intracellular.



Mesostoma ehrenbergii wardii. (Modified from Ruebush, 1940)

Excretion The protonephridia are elongated tubules usually *paired* with ciliated flame bulbs on the lateral branches and one or more external openings. They are often lacking in marine forms and are probably primarily concerned with elimination of water.

Nervous system In primitive forms the nervous system is an epidermal network with five pairs of longitudinal nerves connected by a nerve *ring*. Swellings at the intersections of nerve ring and nerves represent the beginnings of a brain. With further development, a pair of cerebral ganglia are formed by fusion of these swellings and the entire system sinks into the parenchyma. Posteriorly the longitudinal nerves are reduced to one well-developed ventral pair and often a much smaller dorsal pair. Intermediate conditions between these two extremes are common. Sensory receptors are located chiefly in the head region although tactile bristles arise from widely scattered sensory cells. Chemoreceptors consist of depressed areas of epidermis with cilia for circulating water over the sensory surface. These are located in auricles, frontal organs, and ciliated rings, pits, and furrows. Statocysts or organs of equilibrium occur chiefly in primitive marine forms. Many species have one or two pairs of photoreceptors or eyes and in land planarians and polyclads these may become numerous.

Reproductive system and reproduction. Virtually all Turbellaria are hermaphroditic. Male and female systems may be separate throughout or may have a common antrum and pore. Genital pores are

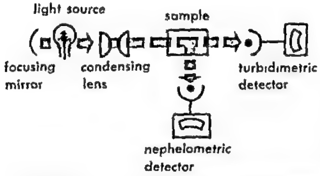
they consist of a few relatively large or several to many small bodies, but some acoeles have only scattered clusters of germ cells in the parenchyma. The male system has a single pair of sperm ducts which may fuse to form a *seminal duct*, enlarge to form spermiducal or storage vesicles, or empty into a true seminal vesicle. The *copulatory organ* is usually muscular, often encloses the seminal vesicle, may be armed with a cuticular apparatus, and contains the ejaculatory duct. When prostate glands are present their secretions may be stored in a prostatic vesicle but are mixed with the sperm in the seminal vesicle or in the *male genital canal*. Some Turbellaria produce entolecithal eggs which contain yolk but generally the eggs are *ectolecithal* and yolk is derived from yolk cells. These yolk cells are probably degenerate oocytes which may be produced throughout the gonad, in a specialized section of the gonad, or in special yolk glands which are distinct from the ovary. The female system usually has its own ducts through which its products reach the female antrum and ultimately the exterior. In the absence of ducts, fertilized eggs escape by rupture of the body wall or by way of the digestive system and the mouth. Accessory structures such as vaginae, copulatory bursae, uteri, seminal receptacles, and specialized glands may also occur. Asexual reproduction by fragmentation or by binary fission with the formation of temporary chains of zooids occurs in some rhabdocoeles and triclads. *See* ACOELA; ALLOEOCOELA; POLYCLADIDA; RHABDOCOELA; TRICLADIDA

[E.R.J.]

Bibliography. E. Bresslau, *Turbellaria*, in W. Kukenthal and T. Krumback (eds.), *Handbuch der Zoologie*, vol. 2, pt. 1, 1933; L. H. Hyman, *The Invertebrates*, vol. 2, 1951.

Turbidimetric analysis

An analytical method based on the measurement of the attenuation of transmitted light by a solution containing a finely divided precipitate. There is a direct relationship between the amount of light attenuated and the amount of material in suspension Either visual comparison with known standards or photoelectric detection is used to measure the amount of light transmitted by the turbid solutions. In the photoelectric method, a standard curve of percent transmittance versus concentration of the material in question is usually prepared and the compositions of unknowns are determined by reference to the curve. The illustration is a schematic diagram of a simple turbidimeter. Turbidity is a function of the concentration and particle-size distribution of turbid material and the path length of the light in the sample.



Simple diagram of a turbidimeter and nephelometer.

Turbidimetry is often not very precise since it is difficult to prepare a stable and reproducible suspension of the precipitate. The amount of light scattered and absorbed by the precipitate also must be neither too great nor too small. A precision and accuracy of ± 5–10% of the amount present is usually obtained, but the method provides a very sensitive means of determining some elements and groups This is illustrated by the determinations of barium or sulfate as barium sulfate, of phosphate as strychnine phosphomolybdate, and silver or chloride as silver chloride. The methods are used to detect a few parts per million of these ions.

In addition to direct measurement or comparison of absorbed light, another common turbidimetric method is the determination of the height of a column of turbid liquid which will obscure a light or an object. The height of liquid is then compared with a solution of known composition or with some sort of standard scale This type of measurement is used in the Jackson candle turbidimeter to measure the turbidity of water and in the visual Parr-sulfur turbidimeter. See NEPHELOMETRIC ANALYSIS; OPTICAL METHODS OF CHEMICAL ANALYSIS

[R.F.C.]

Turbidity current

A submarine flow of sediment or of sediment-laden water which occurs when an unstable mass of sediment at the top of a relatively steep slope is jarred loose and slides down slope. As the slide or slump travels down slope, it becomes more fluid, partly because of the loss of internal cohesive strength and partly because of inmixing of the superadjacent water.

Turbidity currents occur at the edge of the continental slope, in the vicinity of river mouths, and off prominent capes. They are triggered by earthquakes, hurricanes, floods, or simply by the bed load transport of rivers debouching at the edge of continental slopes.

Turbidity currents were first proposed as a hypothesis to account for the erosion of submarine and sublacustrine canyons. Supporting evidence was found in submarine telegraph cable breakage following the Grand Banks earthquake of 1929 and the Orleansville, Algeria, earthquake of 1954 In both cases submarine cables were broken consecutively in the order of increasing distance down slope. All of the cables were broken in at least two widely separated places. The sections between breaks were swept away or buried beneath sediment deposited by the current (Fig. 1).

Sediment cores taken in the suspected area of deposition revealed a recently deposited, graded bed of silt and sand on the abyssal plain Turbidity currents deposit the heaviest grains first and the finest last, and thus their deposits are graded in size from coarse at the base to fine at the top. Graded sands containing shallow-water fossils are found in the beds of submarine canyons, at their mouths, and over the vast abyssal plains of the

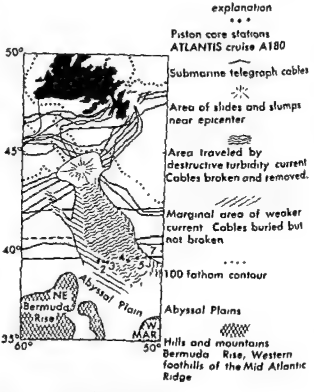


Fig. 1. Area of Grand Banks earthquake of November 18, 1929 Sketch shows areas of cable breakage and cable burial, and relative position of sediment cores and turbidity current

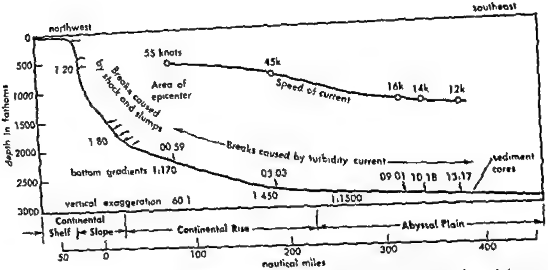


Fig 2 Profile south of the Grand Banks showing progress of the turbidity current Positions of 12 submarine cables are indicated by arrows. For each of the last five, the figure above the arrow indicates the time interval between the earthquake and the arrival of the current that broke the cable Superimposed graph indicates calculated speed of the current as it passed

ocean floor. The turbid ...

yons, and the smoothing of the abyssal plains.

The telegraph cable breaks which occurred after the Grand Banks and Orleansville earthquakes indicated by the time sequence of the breaks that the current attained velocities of at least 50 knots on a slope of 1 50, but slowed to 12 knots on slopes of 1 1500 (Fig 2) Turbidity current breakage of ...

... major rivers, generally during times of flood This implies that the high bed-load transport during flood stage can generate turbidity currents.

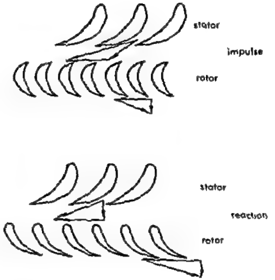
The lowered sea level of the glacial stages must have forced most rivers of the world to empty at or near the top of the continental slope Under those conditions turbidity currents must have been much more important than at present. Turbidity currents undoubtedly transported a large proportion of sediments deposited in ancient geosynclines.

[B C H]

Bibliography H Charnock, Turbidity currents, *Nature*, 183(4662) 657–659, 1959; R A Daly, Origin of submarine canyons *Am J Sci.*, ser 5, 31 401–420, 1936, B. C. Heezen, The origin of submarine canyons *Sci American*, 195(2):36–41, 1956, B C. Heezen and M Ewing, Orleansville earthquake and turbidity currents, *Bull Am Assoc Petrol Geologists*, 39(12)·2505–2514, 1955; B. C. Heezen and M Ewing, Turbidity currents and submarine slumps and the 1929 Grand Banks earthquake, *Am J Sci.*, 250:849–873, 1952; P H. Kuenen, Estimated size of the Grand Banks turbidity current, *Am. J. Sci.*, 250:874–884, 1952; P H. Kuenen and C. I Migliorini, Turbidity currents as a cause of graded bedding, *J. Geol.*, 58·91–126, 1950

Turbine

A *machine* for generating mechanical power in rotary motion from the energy in a stream of fluid. The energy in the fluid, originally in the form of head or pressure energy, is converted to velocity energy in passing through a system of stationary and moving blades in the turbine. Changes in the magnitude and direction of the fluid velocity are made to cause tangential forces on the rotor blades, thus producing power as the rotor turns. The most commonly used fluids are steam (see STEAM TURBINE), hot air or gaseous products of combustion (see GAS TURBINE), and water (see HYDRAULIC TURBINE). Turbines drive about 95% of all electrical



Velocity diagrams and blade sections for impulse and reaction steam turbine stages.

power-producing generators in the world, and are used for innumerable other mechanical drives. *See* TURBINE PROPULSION.

There are two basic design concepts for the blade system (see illustration). In the impulse turbine, the fluid is accelerated through the stationary blades and then made to impinge upon the rotor blades or buckets, thus having its direction changed and producing a force on the rotor with substantially no pressure change while passing through the rotor.

In the reaction turbine, about half of the available energy is used to produce a velocity through the stationary blades approximately equal to the rotor velocity, and the remainder of the pressure energy then accelerates the fluid while it is passing through the rotor, thus producing a driving force from the acceleration. *See* PRIME MOVER [P.H K]

Turbine propulsion

A gas turbine is used to propel a vehicle such as an aircraft or automobile, in contrast to a stationary turbine used in central generating stations and other fixed locations (*see* PROPULSION) It is a form of internal combustion engine (*see* GAS TURBINE).

Basic engine. The gas turbine consists essentially of an air compressor, a combustion chamber, and a turbine wheel (Fig. 1). The compressor and the turbine wheel are on the same shaft and spin together, the assembly being termed a spool. The hot gases from the combustion chamber strike the cupped blades of the turbine wheel at high velocity, driving the shaft and thus rotating the compressor (*see* BRAYTON CYCLE). Air from the atmosphere is sucked into the inlet ports of the compressor and flows progressively through each blade stage (assuming the compressor is of the axial-flow type). As the air pressure increases, its volume decreases until maximum compression is reached at the last blade stage. The highly compressed (and thus high-temperature) air is then discharged into the duct leading to the combustion chamber. This unit has one or more fuel nozzles, through which the fuel is sprayed to mix with the moving air. In starting, the fuel spray is ignited by electric resistance type of ignition plugs. Once ignited, the fuel-air mixture burns continuously, so that the ignition can be switched off.

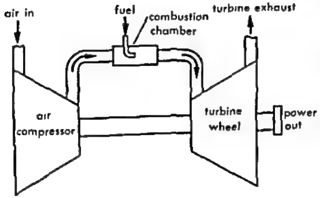


Fig. 1. Simple gas turbine

Instead of the single combustion chamber, which is common to the annular form of chamber employed in many axial-flow compressor gas turbines, a number of separate chambers, known as can type or cannular chambers, are often used. Thus, in a typical aircraft gas turbine, eight to ten can type chambers, each with its own fuel nozzle, are employed.

The products of combustion leave the chamber through a duct and fixed guides or nozzles to enter the turbine unit. Here the gases flow axially between the fixed guides and the rotating blades of the turbine, falling in both temperature and pressure as they deliver most of their energy to the turbine wheel.

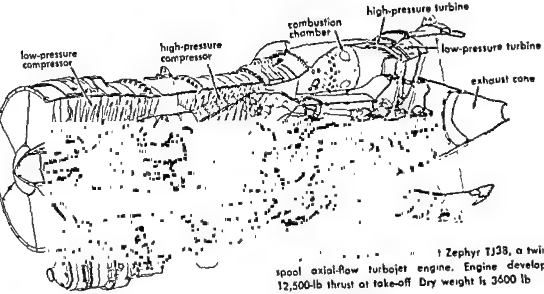
Aircraft applications. The turbine engine has already almost completely replaced the piston engine on military aircraft, and is rapidly invading the commercial field as well, particularly for high speed long-range operations Only for short haul slow-speed transport operations, private flying, executive flying, agricultural operations, and the like does the piston engine still dominate the field. *See* AIRCRAFT ENGINE; AIRCRAFT PROPULSION.

The main reason for the use of the turbine engine is its inherent simplicity. The turbine-compressor unit, with its relatively simple shaft and bearings—usually of the ball or roller type—replaces the complicated piston-engine mechanism for converting the reciprocating motion of the piston to a rotary one. For example, a 12-cylinder V-type reciprocating engine has at least 90 pairs of sliding or bearing members, each a source of friction loss, as compared with four main shaft bearings of the gas turbine.

Other advantages over the piston engine include: (1) higher mechanical efficiency, (2) inherently better balance, (3) no upper limit to size or output, (4) lower weight and smaller frontal area for a given output, (5) ability to operate on less expensive fuels, and (6) simpler lubrication

Turboprop and turbojet engines classified by speed range

Type	Speed range, mph
Turboprop	*0–500*
Single-shaft, single-stage centrifugal flow	
Single-shaft, 2-stage centrifugal flow	
Single-shaft axial flow	
Twin-shaft twin-spool axial flow	
Twin-shaft single-spool axial flow, with free turbine	
Turbojet	500–750
Single-spool axial flow	
Twin-spool axial flow	
Twin-spool axial-flow bypass	
Twin-spool axial-flow bypass, with turbofan	
High pressure ratio turbojet	750–1500+
Single-spool axial flow	
Twin-spool axial flow	
Turbojet with progressively diminishing pressure ratios	1000–1700+
Single-spool axial flow	
High-temperature types	



† Zephyr TJ38, a twin-spool axial-flow turbojet engine. Engine develops 12,500-lb thrust at take-off Dry weight is 3600 lb

The single-shaft aircraft turbine does have certain inherent disadvantages, but these are steadily being overcome or reduced by extensive research and development These are, briefly: (1) high specific fuel consumption, (2) low acceleration or power response, and (3) sensitivity to air inlet temperature

The aircraft turbine engine has developed into two major types the turbojet, and the turboprop Both of these come in a large number of versions, some of which are listed in the table along with their most suitable speed range

Turbojet engine The thrust or push of a turbojet engine comes from reactions set up inside the engine as the gases are accelerated through it and out the exhaust nozzle *See* TURBOJET, TURBORAM JET

The propulsive ability of a turbojet engine is measured in pounds of thrust rather than in horsepower (which is used for piston and turboprop engines) because no actual engine motion is involved when the aircraft is standing still (*see* THRUST) Once the airplane starts to move, however a comparison between thrust and horsepower can be made, with the following equation: $T_{HP} = TS/375$ where T_{HP} is thrust horsepower, T is thrust in lb, and S is speed in mph Thus, at 375 mph, one pound of thrust equals one thrust horsepower

Aircraft turbojet engines range from small units delivering 400–900 lb thrust to huge power plants,

sor is connected by a shaft to the high pressure turbine that drives it Similarly, the low-pressure compressor is connected to the low-pressure turbine that drives it

Many methods have been devised to improve the performance of turbojet engines Two of the most promising are afterburning and the turbofan

Afterburning is a method by which the maximum thrust output of a turbojet engine can be increased, for short periods, by as much as 50% at take-off and several times this amount at high flight speeds

The afterburner takes advantage of the fact that only about 25% of the air passing through the main engine is used for combustion When the afterburner is used, some of this excess air is burned with a large amount of fuel to increase the energy of the gases and thereby the jet velocity and thrust *See* AFTERBURNER.

The turbofan engine is, simply, a jet engine with a fan added In some cases, the fan is mounted in the rear, as in the General Electric CJ805–21. In other engines, the fan is in front. In the engine of Fig 3, the fan is a built-up section of the compressor, which serves somewhat as a propeller, except that it is located inside the engine (The fan of the aft fan engine is driven by exhaust gases aft of the compressor turbine.)

This fan draws in considerably *more air than the* compressor of a conventional jet. But much of the air, after it has been compressed, is released through ducts before it reaches the combustion chambers This bypass air is ducted outside the basic engine, producing additional thrust.

A turbofan engine operates more efficiently, and thus at a lower thrust specific fuel consumption. It also provides shorter take-off distance and lower noise level *See* TURBOFAN.

Turboprop engine. The turboprop engine, also known as the propjet, consists of the basic turbine engine plus a reduction gear through which the turbine drives a propeller to obtain thrust (Fig 4). Some additional thrust is also obtained from the jet exhaust. (The engine may also be designed with shafting suitable for operating helicopters and other shaft-driven vehicles. It is then called a turboshaft engine) *See* HELICOPTER.

The prime advantage of the turboprop arrange-

The upper speed limit of the turboprop engine is

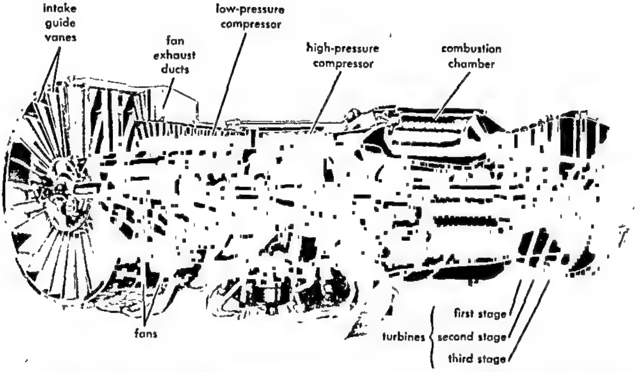


Fig 3. Cutaway view of Pratt & Whitney axial-flow turbofan engine JT3D. Engine has 13 compressor stages, 3 turbine stages, and 2 fan stages. It develops 17,000 lb thrust at take-off. Dry weight is 4025 lb

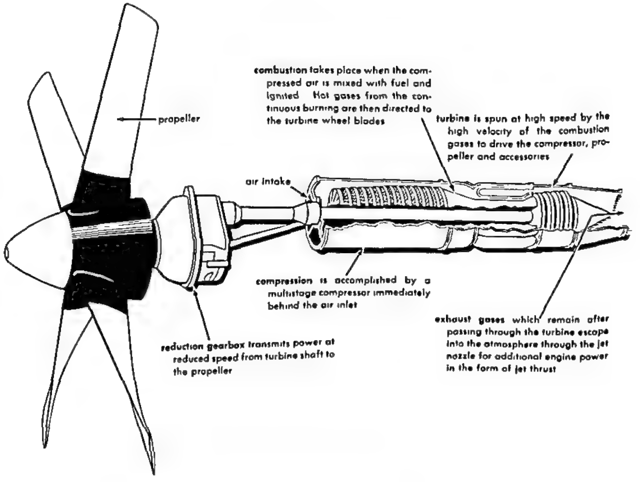
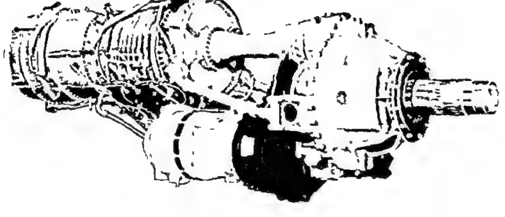


Fig. 4 Main parts of axial-flow turboprop engine and their function (*Allison Div*)



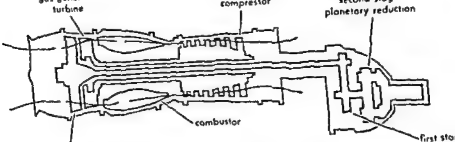


Fig 5 Photograph and diagrammatic sketch of turboprop engine Unit produces 2570 equivalent shaft horsepower, weighs 1079 lb It has 14-stage axial compressor, 2-stage free power turbine It has been designed for helicopters, convertiplanes, troop and supply transports, and antisubmarine and early-warning aircraft (*General Electric Co*)

fixed by propeller efficiency and noise considerations At present the practical limit is about 450 mph but future improvements in high-speed propellers could increase this limiting speed

Some of the ways being used to improve turboprop engines are (1) regeneration, which is the use of the exhaust gas to preheat the compressed air before it reaches the burners, (2) reheating, which is the injection and burning of fuel between stages of the turbine after the first stage, and (3) afterburning, which is also used with turbojet engines, as already discussed

Another modification to the turboprop engine is the use of the free turbine (Fig 5) The propeller is then driven by a separate turbine unit (called the power turbine) supplied with exhaust gases from the compressor driving front turbine unit (called the gas generator turbine). This arrangement gives greater control flexibility, easier starting, better take-off performance, and lower propeller speed

to the inlet of the compressor at the front of the engine. It is compressed by the axial-centrifugal compressor and then delivered through a vaneless diffuser to a collector scroll. The compressed air

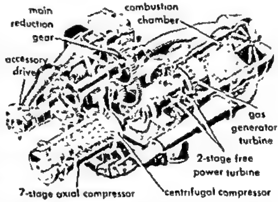


Fig 6 Turboshaft engine produces 250 hp, weighs about 100 lb It has 7-stage axial compressor, one centrifugal compressor, single-stage gas generator turbine, and 2 stage power turbine It can be used on light observation planes, 2-place helicopters, drones, auxiliary power units (*Allison Div*)

then flows through two external tubes (located on either side of the engine at centerline height), which carry the air to the single combustion chamber This, in turn, is located at the rear of the engine coaxial with the compressor turbine shaft. Combustion gases leaving the combustor move forward through the turbine into an exhaust hood in the middle of the engine, where they are collected and exhausted downward.

Special types Two other types of aircraft power plants employing turbines are the turbocompound engine and the nuclear engine.

The turbocompound engine consists of a piston engine plus an exhaust gas driven turbine and a compressor driven by the crankshaft (Fig 7). The turbine provides additional power to the crankshaft, rather than directly driving the compressor (as in a turbocharged engine). The shaft-driven compressor provides the supercharged air for the engine. *See* SUPERCHARGER.

Two examples from the aircraft field are the Wright Turbo Compound engine, which operates on gasoline, and the British Napier Nomad, which is a diesel engine plus turbine. In the gasoline engine, three exhaust gas driven turbines are coupled to the crankshaft through reduction gearing and fluid couplings In both engines, the added power from the turbine results in lower specific weight and better specific fuel consumption when cruising (Another type of compound engine—the free-piston-turbine engine—is discussed under automotive applications because it has not been tried for aircraft use.)

In one form of nuclear power plant, energy from an atomic reactor is transferred to a turbine for power utilization Work in this field is progressing The chief problem is the high radiation level at the reactor center, which requires extensive shielding and the development of radiation resistant materials for use within the shielded volume *See* NUCLEAR AIRCRAFT PROPULSION

Automotive applications. The gas turbine, either alone or in combination with a piston engine (resulting in a compound engine, such as the free-

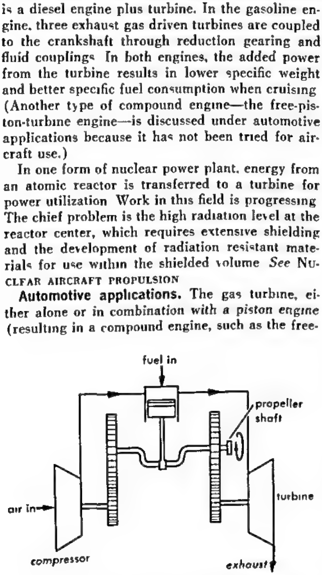


Fig 7. Turbocompound engine consists of piston engine plus exhaust gas driven turbine and shaft-driven compressor.

piston-turbine engine), is being developed for application to various ground vehicles. Many car companies, both here and abroad, are carrying on extensive experimental work and some of them have produced cars, trucks, buses, or tractors with experimental turbine engines Gas turbines are used commercially in railroad engines but only to an extremely limited extent.

There are advantages for gas turbines in ground vehicles, but there are also some difficulties. Some of the advantages are: (1) *transmission* can be simplified because the gas turbine comes equipped with built-in torque converter characteristics, (2) it is inherently simple, (3) it is of small size and low weight, (4) it has an ability to burn a broad range of fuels, (5) it can operate in any climate, and (6) it has low maintenance

Despite these advantages, progress is slow for several reasons The piston engine keeps advancing so that it is increasingly difficult for other types to replace it (*see* AUTOMOTIVE ENGINE). Projected higher compression ratios, greater use of aluminum, and new developments in fuel injection, for example, will all result in the improvement of gasoline engine performance Similarly, such developments as a greater degree of turbocharging and improved methods of diesel combustion are bringing increased performance to the diesel engine.

In addition, there are still some difficult turbine problems to be solved. Ways must be devised, for example, to reduce: (1) fuel consumption, (2) acceleration delay, and (3) costs.

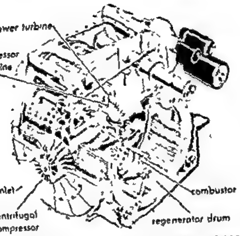
To reduce fuel consumption, engineers are using such accessories as intercoolers and regenerators.

One way of coping with acceleration lag is to provide a wastegate power control between the gas generator turbine and the power turbine When open, this allows the exhaust gas to bypass the power turbine, while the gas generator turbine is turning at rated speed This would be the condition required when the vehicle is stationary When the wastegate is closed, the gas is directed against the power turbine and torque response is instantaneous

Satisfactory low-cost noncritical materials and production techniques are being developed to reduce costs Of course, high-volume production itself will bring costs down, as compared to those of experimental units Recent developments have been made in the straight gas turbine engine and the free-piston engine compounded with the turbine.

Turbine engine The first gas turbine driven automobile was demonstrated by Rover in England in 1950 In this country, Ford, General Motors, and Chrysler all have experimental gas turbine cars, while Boeing Aircraft has developed a 240-hp gas turbine engine that has been successfully operated in a Kenworth truck. One type of turbine unit uses regeneration to recover heat from the turbine exhaust (Fig. 8).

Free-piston engine The free-piston engine combines a free-piston gasifier, which is essentially an air compressor run by a diesel engine with a power



8 Cutaway view of General Motors GT-305 irlfite engine, rated at 225 hp Dry weight in- ding all accessories is 590 lb

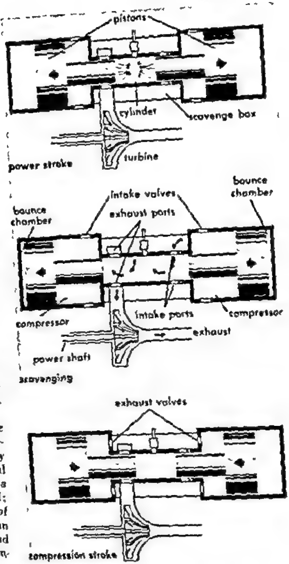


Fig 9 Operation of free-piston engine.

rbine which is run by the exhaust from the gasi- r. Interest in the *free-piston engine* for operation vehicles is relatively recent In this country, Ford nd General Motors are actively engaged in its de- elopment for possible use in passenger cars, rucks and tractors

The reasons for this development stem from the ollowing advantages (1) less alloy content is re- quired than in the gas turbine, (2) fuel economy s slightly better than that of the conventional 9.5-1 piston engine and much better than that of a regenerative gas turbine, particularly at part load; (3) throttle response is about the same as that of a piston engine with torque converter; (4) it can run on a wide range of fuels, and (5) gasifier and power turbine can be located quite far apart, con- nected only by hot gas ducts

There are, however, problems to be solved- (1) its weight and bulk are too high, (2) to give

celeration lag, although to a lesser extent than with the straight gas turbine engine, and (6) there are operational problems of starting idling, control, and regulation *See* DIESEL CYCLE, TURBINE

Figure 9 shows diagrammatically the operation of the free piston engine On the power stroke the pistons move outwardly in the direction of the arrows due to combustion in the central diesel cylinder The pistons uncover the exhaust ports first, allowing blowdown of cylinder pressure to the turbine As the intake ports are opened, air previously compressed into the scavenge box blows through the cylinder, out the exhaust ports, and to the turbine During this stroke, air is also drawn into the compressor through the intake valves During the power stroke air has been compressed in the bounce chambers which stops the pistons and forces them inward

On the compression stroke the ports are closed, starting the diesel compression process The compressor-cylinder intake valves have closed and air is forced into the central scavenge box through the compressor exhaust valves When the pistons reach the innermost position, fuel is injected into the diesel cylinder, starting combustion, and the cycle repeats.

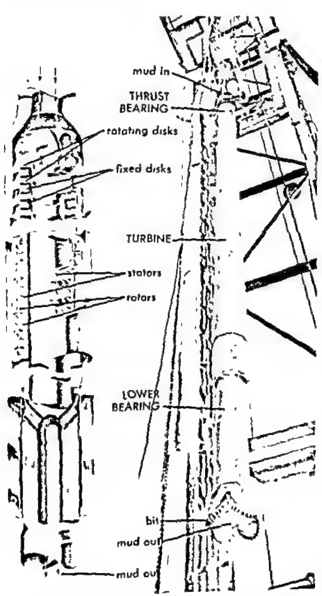
A free-piston gasifier has been linked to a power turbine and installed in a farm tractor The engine develops 50–150 hp, is 37 in. long and 16 in in diameter In tests the engine has operated smoothly and is practically vibration-free. Mufflers in the exhaust line are not necessary, since the noise level is low A turbine whine is perceptible at idle speeds, but it is not audible at working speeds With all accessories and an auxiliary gear box this free-piston power plant weights 1125 lb.

This installation was made because it was considered to be ideal as a farm tractor power plant

sel piston engine. It has the response and flexibility of a gasoline engine. An auxiliary starting engine is not required, as is generally necessary with diesels. A wide range of engine sizes can be economically produced from the same manufactured parts. It is adaptable to a wide range of fuels and has excellent torque characteristics. [F.AL.]

Turbodrill

A rotary tool used in drilling oil or gas wells in which the bit is rotated by a turbine motor inside the well. The principal difference between rotary and turbodrilling lies in the manner power is applied to the rotating bit, or cutting tool In the rotary method, the bit is attached to a drill pipe which is rotated through power supplied on the surface In the turbodrill method, power is generated at the bottom of the hole by means of a mud-operated turbine. *See* ROTARY TOOL DRILL.



Bit of turbodrill is rotated by a turbine built inside drill pipe and powered by mud pumped into well to cool bit. Turbine is held and lowered into well by drill pipe (Dresser Industries)

The turbodrill (see illustration) consists of four basic components: the upper, or thrust, bearing; the turbine; the lower bearing; and the bit Most turbodrills are about 30 ft long, with shafts about 20 ft long. The turbodrill is attached at its top to a drill collar, or heavy length of steel pipe, that makes up the bottom end of the drill pipe extending to the surface. Once the turbodrill passes below the well head, operations on the rig floor are the same as for rotary drilling. Rotation of the drill pipe is not necessary for turbodrilling, because rotation of the bit develops through the turbine on the lower end of the drill string. It is usual practice, however, to rotate the drill pipe above the turbine slowly, at from 6–8 rpm, either by means of the rotary table on the derrick floor, or through torque of the turbine on the bottom. Rotation of the bit is much faster than in rotary drilling, and is usually between 500 rpm and 1000 rpm.

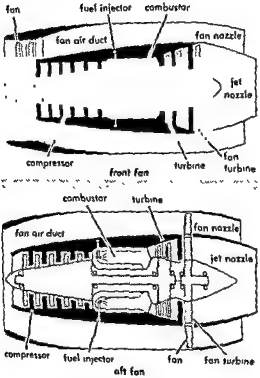
In operation, mud is pumped through the drill pipe, passing through the thrust bearing and into the turbine In the turbine, stators attached to the body of the tool divert the mud flow onto rotors attached to the shaft. This causes the shaft, which is connected to the bit, to rotate. The mud passes through a hollow part of the shaft in the lower bearing and through the bit, as in rotary drilling to remove cuttings, cool the bit, and perform the other functions of drilling fluid. Capacity of the mud pump, which is the power source, determines rotational speed.

Two basic types of turbodrills are used in the United States One is a standard 100-stage unit (one rotor and one stator comprise a stage); the other is a tandem turbodrill, made up of two or three standard sections Although turbodrills have been in wide use in the Soviet Union for several years, they are still relatively rare in the United States, where the emphasis has been on the rotary tool method of drilling Despite faster penetration with the turbodrill, and several other advantages, widespread use of the turbodrill in the United States has been limited principally because of the faster wear of the bits, necessitating time consuming and costly round trips to remove the drill string from the hole and change the bit *See* CABLE-TOOL DRILL, OIL AND GAS WELL DRILLING [A.L.P.]

Turbofan

An air-breathing jet engine with operational characteristics between those of the turbojet and the turboprop Like the turboprop, the turbofan consists of a compressor-combustor-turbine unit (called gas generator) and a power turbine which in this case drives a low-pressure ratio compressor called a fan

Operating principle. The gas generator produces useful energy in the form of hot gas under pressure Part of this energy is converted through the turbine and the fan that it drives into kinetic pressure of the fan airflow, which is expanded in a fan air nozzle, thereby being converted into kinetic en



Turbofan engine configurations.

ergy. The rest of the gas generator energy is converted into kinetic energy through expansion in a jet nozzle. In this manner the turbofan produces useful thrust through two separate streams, the gas generator flow and the fan flow.

There is an optimum energy split between both streams which is a function of component efficiencies, temperature ratios and flight speed. Both streams may be separately discharged through concentric jet nozzles or mixed before expansion in a common jet nozzle. Mixing the mass flows improves the efficiency slightly and reduces the noise level.

Design characteristics. The ratio of fan mass flow to gas generator mass flow is called the *bypass ratio*. Bypass ratios are generally 0.5–2.5, depending upon the desired operational characteristics.

The turbofan cycle is inherently superior in thrust characteristics and propulsion efficiency to the turbojet cycle for subsonic flight because it produces the desired thrust by accelerating a larger mass of air by a smaller velocity increment. Compared to the turboprop it offers higher speed capability, smaller over-all diameter, less noise, and less mechanical complexity. For these reasons there is a definite trend toward use of the turbofan engine as a power plant for subsonic jet-type aircraft. See TURBOJET; TURBOPROP.

High pressure ratios in the gas generator and high turbine-inlet temperatures tend to improve the over-all efficiency of the turbofan; therefore, it shares gas generator development problems with the turbojet and the turboprop engine.

Configurations. A variety of turbofan configurations is possible. It is always desirable to keep the fan rotor mechanically separate from the gas generator rotor because independently variable rotational speeds lead to highest efficiency. The diagram illustrates a front fan arrangement and an aft fan configuration, which is particularly attractive if an existing turbojet engine is to be converted into a turbofan. Both designs have their specific advantages.

Large bypass ratio turbofans with geared low-speed fans are excellent power plants for smaller subsonic jet aircraft with speeds under 400 mph because of their excellent thrust-weight ratio. Fans with high bypass ratios will be utilized to produce vertical lift in VTOL aircraft, designed to rise and descend vertically without extensive runways. See VERTICAL TAKE OFF AND LANDING (VTOL).

[P.G.KA.]

Bibliography: J. V. Casamassa and R. D. Bent, *Jet Aircraft Power Systems*, 2d ed., 1957; A. W. Judge, *Gas Turbines for Aircraft*, 1958.

Turbojet

A propulsion engine used to power most military fighters and bombers, and some transports. Commercial airlines are beginning to use turbojet-powered aircraft.

Operating principle. In a turbojet, as illustrated, air approaches the inlet diffuser at a relative velocity equal to the flight speed. In passing through the diffuser the velocity of the air is decreased and its pressure is increased (see DIFFUSER). The air pressure is increased further as it passes through the compressor. In the combustion chamber a steady stream of fuel is injected into the air and combustion takes place continuously. The high pressure hot gas passes through the turbine nozzles, which direct it at high velocity against the buckets on the turbine wheel, thereby causing the wheel to rotate. The turbine wheel drives the compressor to which it is connected through a shaft. This is the sole function of the turbine.

After the hot gas leaves the turbine, it is still at a high temperature and at a pressure considerably above atmospheric. The hot gas is discharged rearward through the exhaust nozzle of the engine at a high velocity (see BRAYTON CYCLE).

The thrust obtained is equal to the over-all increase in momentum of the gas as a result of its passage through the engine. This thrust is

$$F = M(V_j - V_a)$$

where M is the mass flow of gas per second through the engine, V_j is the exhaust jet velocity, and V_a is the airplane velocity.

Compressors. Two types of compressors are used on turbojet engines—centrifugal-flow and axial-flow compressors. The centrifugal compressor is the simpler of the two and was used on the early versions of the engine (see COMPRESSOR). For example, it was used on the f

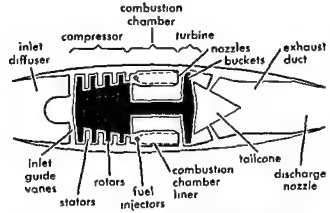


Diagram of an axial-flow subsonic turbojet engine.

signed by Frank Whittle, which was the forerunner of a series of early American engines, namely, the I-A, I-16, and I-40 (which became the J-33).

The trend has been consistently toward the application of the axial-flow compressor because of its greater efficiency and greater air-handling capacity per unit frontal area, in spite of its complexity and fragility. Engines such as the J-57, the J-75, and the J-79 are of this type

Compressor stall. One of the problems associated with the axial-flow compressor is that of designing the stator vanes to direct the air into the rotor vanes at the proper angle at all rotational speeds If the air is directed at a vane at too sharp an angle, it will not follow the vane but will break away. Loss in pressure and efficiency occurs and even possibly vibration of the blade.

Axial-flow compressors are designed for the full power condition At engine rotative speeds below 70% of maximum, the compressors on some engines develop a rotating stall One or more stall cells or areas of stall form around the compressor, in these cells the flow strikes the blades at an improper angle of attack. The stalled areas move around the compressor, and each time one passes a given blade, the blade is subject to an impact. If these impacts come in resonance with a natural frequency of the blade, severe and sometimes destructive vibration may result.

These problems of low efficiency and stall at low rotative speeds are particularly severe in compressors with many stages. One solution is to separate the compressor into two sections, each mounted on a separate shaft coaxial with the other and driven by separate turbines An engine of this type, called a two-spool engine, is the J-57. Another solution to these problems is to provide a mechanism for adjusting the angular setting of the stator vanes with change in rotative speed. The J-79 engine uses this method. Interstage bleedoff of air at low rotative speeds is also used to assist in combating stall on some engines as in the J 57. Other schemes have also been investigated.

Foreign object damage. ... is problem has been the ingestion ... by the engine. A foreign object ... compressor may cause blad... a nick in a blade, which can become the nucleus of a fatigue failure. The failure of one blade can set off an avalanche of failures as it passes through the compressor. Compressor failures have been caused by objects picked up from the external environment, by objects shaken loose from the engine, or by objects left by mechanics. These latter two sources can be eliminated by careful design and maintenance.

Foreign objects have been sucked up from runways during take-off and landing with the assistance of vortices generated by the engine. Forward directed air jets, which operate during take-off and break up these vortices, have been suggested Screens are also used to stop the larger particles However, particles smaller than the screen mesh can get through Screens have the disadvantages that their resistance to the air flow imposes a loss in performance, and they provide a surface to which ice can adhere in an icing atmosphere For this reason, they are designed to be retracted once the airplane is aloft. By keeping runways clean and by application of these various aids, the possibility of foreign object damage can be greatly reduced

Engine control. The elements that can be acted upon to control engine operation are the fuel throttle and, for some engines, the variable-angle compressor stator vanes The fuel flow is increased during engine start-up as the rotational speed increases The fuel flow is decreased as altitude is increased, to compensate for the reduction in air density.

The following are some conditions encountered in turbojet operations which are associated with engine control:

1 Overtemperature of the combustion gas during start-up can damage turbine buckets.

2. Overspeed of the turbine and overtemperature of the combustion gas at maximum thrust can overstress turbine wheels and buckets

3. Compressor surge during acceleration of the engine may cause damaging vibration of compressor blades

4 Flameout (cessation of combustion) can occur when engine speed is reduced at high altitude.

5 Flameout and also compressor surge can occur when an engine is jockeyed in the course of landing the airplane

Automatic engine controls are designed to relieve the pilot of the task of avoiding these undesirable operating conditions The earliest automatic controls incorporated only a maximum rotative speed governor and an ambient-pressure sensing element that adjusted fuel flow for change in altitude. The trend has been toward more sophisticated automatic controls that handle a greater number of the control requirements at the expense of greater complexity The perfect automatic control has yet to be developed Failure of the control and fuel system represents an important cause of engine failure in flight.

Typical performance. For the ... of illus... g the performance of ...

Maximum continuous rating of Pratt and Whitney J 57 engine

Flight speed, knots	Sea level		25,000 ft		45,000 ft	
	Thrust, lb	Specific fuel consumption, lb/(hr)(lb)	Thrust, lb	Specific fuel consumption, lb/(hr)(lb)	Thrust, lb	Specific fuel consumption, lb/(hr)(lb)
0	9500	0 76				
200	8000	0 87	4700	0 82	2005	0 85
400	7650	0 99	4750	0 92	2250	0 93
600	7600	1 07	5000	0 98	2500	0 98

maximum continuous performance of an engine is displayed in the table. *See* PROPULSION; SPECIFIC FUEL CONSUMPTION. [B PI]

Turboprop

A gas turbine powerplant producing shaft power for aircraft using a propeller The turboprop engine has basic components similar to those of a *turbojet:* compressor, combustor, and turbine In addition, it has a power turbine. This power turbine extracts usable shaft power from the engine mass flow and, drives a conventional propeller

engine rotor hence this type is commonly called a fixed turbine engine This has certain disadvantages in regard to starting, partial load fuel consumption, and windmilling drag.

Figure 2 shows a more advanced turboprop type Here the power turbine has no mechanical connection with the main engine rotor and consequently operates at a different rotational speed, this type being called a free-turbine engine. This arrangement improves the operating characteristics and, especially, the cruise efficiency under partial-load conditions This form of engine is also used to drive helicopter rotors and is then called a gas turbine.

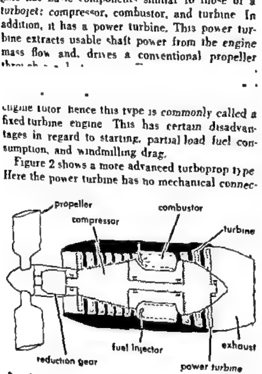


Fig 1 Turboprop engine with connected power turbine

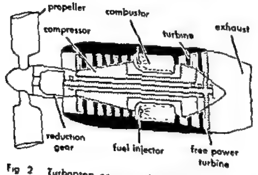


Fig 2 Turboprop engine with free power turbine

Performance characteristics. Compared to the *turbojet* and, to a lesser degree, to the turbofan, the turboprop offers lower fuel consumption and a higher takeoff thrust. It has low engine noise level. Its propellers can be reversed to shorten the landing run For these reasons the turboprop is an excellent power plant for aircraft where these qualities are important. Its disadvantages are heavier weight and increased complexity and maintenance cost Because of propeller characteristics, the turboprop usually reaches peak operating efficiency at lower cruise speeds than the turbofan and is, therefore, better suited for transports in the speed range below 450 mph, although it is basically possible to reach high subsonic and even supersonic flight speeds with a turboprop At high altitudes the turboprop achieves lower cruise fuel consumption levels than those of the best reciprocating engines (approximately 0 34 lb of fuel per equivalent shaft horsepower-hour).

Design considerations. For maximum efficiency the turboprop engine should have a high pressure ratio (between 8 and 14) and high turbine-inlet temperatures (1800–1900°F) Its specific fuel consumption decreases with rising turbine-inlet temperatures, in contrast to the turbojet cycle.

The design limitations of the turboprop are similar to those of the turbojet, and it is subject to similar problems *See* TURBOJET.

Its high pressure ratio leads to compressor stall under adverse operating conditions and requires appropriate measures for control such as interstage bleed, variable stators, or a twin spool rotor *See* TURBOFAN

Control problems The turboprop has the same control variables as the turbojet, as well as a number of additional ones These include fuel flow, affecting rotor speed, and variable stator position or bleed valve position, depending upon the system selected. In addition, there is the propeller pitch control which, with fuel flow, affects the speed of the power turbine. It is generally desirable to maintain gas generator speed

to select the desired horsepower level through proper adjustment of fuel flow and propeller pitch. The free power turbine configuration allows propeller rotational speed to be an additional variable.

Compensation for altitude and ram air density is required as in the turbojet engine.

Regeneration. The efficiency of a turboprop engine can be increased through use of a regenerator, especially on engines with only a moderate pressure ratio (*see* BRAYTON CYCLE). Early designs attempted to incorporate this feature. Advances in compressor technology made it possible to obtain almost equivalent fuel consumption levels through high pressure ratios and high turbine-inlet temperatures at lower over-all complexity and weight, so that regenerators are not widely used in modern turboprop engines. [P.G.KA.]

Bibliography: V. C. Finch, *Jet Propulsion-Turboprops*, 1955; J. G. Keenan, *Elementary Theory of Gas Turbines and Jet Propulsion*, 1946, A. W. Morley, *Aircraft Propulsion*, 1953

Turboram jet

An aircraft engine in which intake air is compressed at high flight speeds by the ram effect of the intake moving into the air ahead of the engine and at low flight speeds by the compressor rotating behind the intake, and in which propulsion is produced by a high-speed exhaust For operation from standstill to high flight Mach numbers, the turboram jet engine provides useful performance characteristics The compressor is driven by a turbine for engine operation at low flight speeds (*see* TURBOJET). At high supersonic flight speeds, the ram entry of air into the engine makes the compressor and its turbine superfluous (*see* RAMJET) Under this operating condition the fuel is burned to greater advantage in a combustion chamber *downstream* of the turbine instead of between compressor and turbine as for the turbojet mode of operation. Also, this second combustion chamber after the turbine can provide additional thrust at low-speed as a turbojet (*see* AFTERBURNER)

Turbojet with afterburning. The turbojet with afterburning is the simplest of the class of engines called turboram jets. The chamber for afterburning is provided with fuel injection nozzles and flame holders (Fig. 1). The temperature of the gas out of the primary combustion chamber is usually limited to about 1700°F to avoid weakening the blades of the turbine. This limits the fuel burned in the primary burner to an amount that consumes about one-third of the available oxygen. The remaining oxygen is used for combustion of the fuel in the secondary combustion chamber (afterburner). A typical temperature of the gas out of the turbine in subsonic flight is in the order of 1300°F (1760°R). This temperature can be raised to about 3000°F (3460°R) in the afterburner. If the pressure drop in the secondary burner is negligible, then, to a first approximation, exhaust jet velocity V_j is proportional to the square root of the temperature of the gas approaching the exhaust nozzle. Hence

$$\frac{V_{jab}(\text{with afterburning})}{V_j(\text{without afterburning})} = \sqrt{\frac{3460}{1760}} = 1.40 \quad (1)$$

The thrust F (lb) of the engine to a first approximation is given by

$$F = M(V_j - V_0) \quad (2)$$

where V_0 is the flight speed (ft/sec) and M is the mass flow rate of gas passing through the engine (slugs/sec) (One slug equals 32 2 lb.) The mass flow of gas through a given engine with and without afterburning is the same Hence, the thrust, with and without afterburning, is

$$\frac{F(\text{with afterburning})}{F(\text{without afterburning})} = \frac{V_{jab} - V_0}{V_j - V_0} \quad (3)$$

At take-off, V_0 is close to zero, thus afterburning has multiplied the thrust for take-off by a factor of 1 4.

Effect of afterburning. The thrust augmentation by afterburning increases substantially with increase in flight speed. For example, at a flight

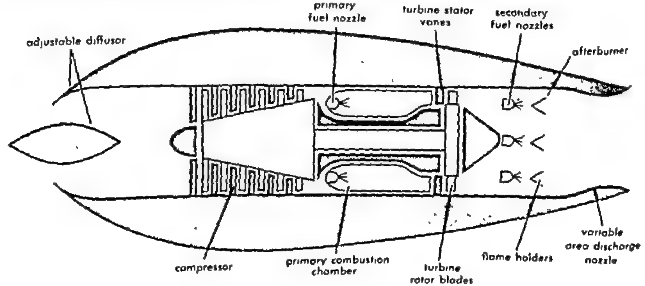


Fig. 1 Turbojet engine with afterburner.

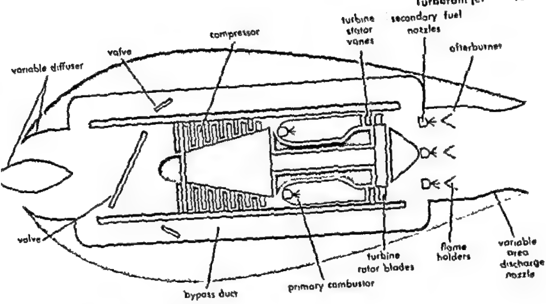


Fig 2 Schematic diagram of dual cycle engine

Mach number of 2 above 36 000 ft altitude, the value of V_0 is about 2000 ft sec, the turbine outlet temperature is about 1500°R and jet velocity V_j without afterburning is about 3000 ft/sec Under these conditions, afterburning to an exhaust temperature of 3460°R provides an increase in jet velocity of

$$\frac{V_{jab}}{V_j} = \sqrt{\frac{3460}{1500}} = 1.5 \qquad (4)$$

and the jet velocity V_{jab} is then $1.5 \times 3000 = 4500$ ft sec The thrust augmentation from Eq (3) is

$$\frac{F(\text{with afterburner})}{F(\text{without afterburner})} = \frac{4500 - 2000}{3000 - 2000} = 2.5$$

For this case with afterburning the thrust is 2.5 times as great as that without afterburning The total fuel consumption with afterburning is roughly three times as great as that without it As a consequence specific fuel consumption is $3/2.5 = 1.2$ times as great In spite of this disadvantage in higher specific f...
engine
augmen
a Mach
lower s... engine weight and specific frontal area, although at Mach 2 the range of an airplane is about the same with or without afterburning (see AIRCRAFT PROPULSION) As flight Mach number increases the disadvantage in specific fuel consumption of the afterburning engine tends to reduce At Mach 3–4 and higher the afterburner engine is a more efficient type than the simple turbojet

The afterburning engine must be provided with a variable area discharge nozzle (Fig 1) because of the large difference in discharge area required with and without afterburning and because of the area adjustment required with large variation in flight Mach number. For most efficient operation an engine required to operate at supersonic flight speed should also have an adjustable inlet diffuser.

Engine controls for the secondary fuel flow, the variable area nozzle, and the inlet diffuser are integrated with the primary engine control ...

... one of the difficult design problems associated with this type of engine Another problem that usually imposes considerable development testing is the design of the flame holder and fuel injection system This system should provide efficient combustion over the required range of afterburner fuel flow rates and flight altitudes, without introducing excessive pressure drop in the flow through the afterburner

Dual cycle engines. Over-all engine pressure ratio is the ratio of the pressure in the combustion chamber to the ambient air pressure This over-all pressure ratio is a major factor in determining the efficiency of such engines as the turbojet, ram jet and turboram jet A high pressure ratio has associated with it a large increase in compression temperature In the turbojet engine the combustion discharge temperature must be limited to a value that the turbine can withstand. The compressor discharge temperature must be considerably lower than this value to allow an appreciable amount of fuel to be burned in the primary combustor. These considerations place an upper limit on the over-all pressure ratio desired in a turbojet engine.

As flight Mach number is increased, the pressure ratio across the inlet diffuser of the turbojet engine increases and the pressure ratio provided by the engine compressor must therefore be decreased In addition, ram compression raises the temperature of the air entering the compressor. The pressure ratios provided by a typical inlet diffuser at flight Mach number of 3 and 4 are 27 and 73, respectively As flight Mach number increases through this range, it is desirable to eliminate the compres-

sor and turbine from the path of the propulsion gas and thus convert the engine to a ram jet

Unfortunately, the elimination of the turbojet engine from the system involves complications in ducting and valving. One method of accomplishing this is to provide ducts from the diffusor to the afterburner that bypass the turbojet engine (Fig 2). Valves are provided in the bypass flow system and the turbojet engine system to divert the air flow from one system to the other.

Because of the importance for supersonic aircraft of drag minimization, low engine frontal area is essential This poses the problem in the system shown in Fig. 2 of designing a compact installation without introducing excessive loss in engine efficiency *because of sharp bends and restrictions in the bypass ducts.* The attainment of high reliability in the valves, valve activators, and controls, is another requirement.

Simply to stop the flow of fuel to the primary burners when ramjet operation is desired and allow the turbine rotor to windmill is ineffective. [B PT.]

Turbulent flow

Motion of fluids in which local velocities and pressures fluctuate irregularly Most flows observed in nature, such as rivers and winds, are turbulent. Such flows occur at high Reynolds numbers In turbulent flow, motion of the fluids is steady only in so far as the temporal mean values of velocities and pressures are concerned The velocity and the pressure distributions in turbulent flows as well as the energy losses are determined mainly by the turbulent fluctuations.

Random nature. The essential characteristic of *turbulent flow is that the fluctuations are random.* Hence, the solution of the turbulence problem re*quires the application of methods of statistical* mechanics. In turbulent flow, the most important phenomenon is the transfer of forces by such random motion The rate of turbulent transfer such as heat transfer, shearing stress, and diffusion, is much higher than that due to molecular mechanism in laminar flow

In turbulent motion, even though the fluid is regarded as a continuum with an average over-all molecular motion, turbulent velocity fluctuations must be superimposed on the mean motion The separation of mean motion and turbulent fluctuation depends mainly on the scale of turbulence Different scales give different descriptions of turbulent flow. Once the scale of turbulence is chosen, the instantaneous velocity component u_i, for instance, is

$$u_i = \bar{u}_i + u'_i \tag{1}$$

where u_i is the *i*th component of the total fluid velocity, $\bar{u}_i$ is the *i*th mean velocity component and u'_i is the *i*th component of the turbulent fluctuating velocity.

Turbulent stresses. The instantaneous velocity component u_i satisfies the Navier-Stokes equations of motion of a viscous fluid. The substitution of the expression for the instantaneous velocity components into the Navier-Stokes equations and the use of the mean values of the equations give the Reynolds equations for turbulent flow. The difference of the Reynolds equation from the Navier-Stokes equations is due to the additional terms of turbulent stresses. The turbulent normal stresses are $-\rho\overline{u_i'^2}$, and the turbulent shearing stresses are $-\rho\overline{u'_iu'_j}$, where $i \neq j$, and ρ is the density of the fluid. These stresses represent the rate of transfer of momentum across the corresponding surfaces because of turbulent velocity fluctuations

Semiempirical theories of turbulence. To illustrate various semiempirical theories of turbulent flow, consider a simple parallel mean flow $\bar{u} = \bar{u}(y)$, $\bar{v} = 0$, $\bar{w} = 0$, with fluctuating velocity components u', v', and w', where u, v, and w are the x, y, and z components of velocity, respectively Semiempirical theories of turbulence are formulated based on various hypotheses about the turbulent stresses

Boussinesq introduced the turbulent exchange coefficient ϵ such that

$$\overline{u'v'} = -\epsilon\frac{d\bar{u}}{dy} \tag{2}$$

In actual analysis, further hypotheses are necessary about the variations of ϵ, which are different for different flow

L. Prandtl originated the mixing-length theory in which the fluctuating value of a transferable quantity q of the fluid may be written as

$$|q'| = l\frac{d\bar{q}}{dy} \tag{3}$$

where l is the mixing length. For simple parallel flow, the exchange coefficient becomes

$$\epsilon = -\overline{lv'} \tag{4}$$

There are many different mixing-length theories based on different quantities being transferred In Prandtl's momentum transfer theory, the momentum of the fluid elements is assumed to be preserved in the mixing process. Prandtl obtained his famous formula for shearing stress τ of nearly parallel turbulent flow as

$$\tau = \rho l^2\left|\frac{d\bar{u}}{dy}\right|\frac{d\bar{u}}{dy} \tag{5}$$

This formula has been used successfully in the calculation of many turbulent flow problems such as flow along a flat plate and in jet and wakes.

In Taylor's vorticity transfer theory, the transferable quantity is vorticity Theodor von Kármán made a similarity hypothesis to determine the mixing length so that no special model for a transferable quantity is required.

Logarithmic velocity profile. The most important deduction from von Kármán's similarity hypothesis is the universal velocity distribution for the flow in circular pipes or between parallel plane walls. Von Kármán first pointed the ratio

tween the velocity defect $U_m - \bar{u}$ and the quantity $\sqrt{\tau_0/\rho}$ is a universal function of the ratio $(y_0 - y)/y_0$, that is

$$\frac{U_m - \bar{u}}{\sqrt{\tau_0/\rho}} = f\left(\frac{y_0 - y}{y_0}\right) \tag{6}$$

here y_0 is the radius of the circular pipe or the half width between the two plates, y is the distance from the wall, $\bar{u}$ is axial velocity, U_m is the maximum axial velocity occurring at $y = y_0$, which is the center of the channel, and τ_0 is the shearing stress at the wall $y = 0$

The universal velocity distribution is

$$\bar{u} = 2.5\sqrt{\frac{\tau_0}{\rho}}\log\frac{y + \Delta}{\delta} \tag{7}$$

where Δ is in the same order of magnitude as the thickness of the laminar sublayer, which is negligible. For smooth wall, length δ is determined by a physical parameter such as density and viscosity of the fluid, for rough surface, it is determined by the roughness of the wall.

For engineering application, a nondimensional pipe resistance coefficient λ is used such that

$$\frac{dp}{dx} = \frac{p_1 - p_2}{L} = \frac{\lambda}{d}\frac{\rho}{2}u_0^2 \tag{8}$$

where $p_1 - p_2$ is the pressure drop along a pipe of length L and diameter d, and u_0 is the average mean velocity over a section of the pipe.

For smooth pipe,

$$\frac{1}{\sqrt{\lambda}} = 2.0\log_{10}\left(\frac{u_0 d}{\nu}\sqrt{\lambda}\right) - 0.80 \tag{9}$$

For pipe of rough surface, the resistance depends on size, shape and spacing of the roughness elements. Only for closely packed roughness, can linear dimension h of the roughness alone be used to describe the roughness. For a completely rough pipe in which $(\sqrt{\tau_0/\rho})h/\nu > 100$, where ν is the coefficient of kinematic viscosity, the resistance law is

$$\frac{1}{\sqrt{\lambda}} = 2.00\log_{10}\left(\frac{y_0}{h}\right) + 1.74 \tag{10}$$

In the noncircular pipe the characteristic length is often represented by the hydraulic mean length L_h, which is $L_h = 2A/L_w$, where A is the cross-sectional area of the pipe, and L_w is the wetted circumferential length. If the hydraulic mean length is used instead of the radius of the circular pipe resistance law (9) may be used for noncircular pipes with an accuracy within a few per cent.

Turbulent jet mixing. Another type of turbulent flow without a solid wall in the flow field is known as the free-turbulence problem. Jet mixing and wakes fall in this category (see WAKE FLOW).

In free-turbulence problems, the application of Prandtl's mixing length theory is more successful than that for the turbulent boundary layer flow along a solid wall. In the free-turbulence problem, simple and plausible assumptions on the variation of mixing length in the flow field are possible. The mean velocity distribution calculated on the basis of these assumptions agrees well with the experimental results over a major portion of the flow field.

For a turbulent jet in a medium at rest, the jet spreads linearly. For a wake of a body of revolution, the width of the wake increases with $(C_D S_b x)^{1/3}$ where C_D is the drag coefficient of the body, S_b is the reference area of the body, and x is the distance from the body.

For a first approximation, velocity distributions in a jet mixing region and in a wake may be represented by error functions.

For turbulent jet mixing of fluids of different temperatures or of different densities, the spread of temperature and of concentration are about the same, and they are usually wider than the spread of velocity profile.

Statistical theory of turbulence. Even though the semiempirical theory had successfully predicted mean velocity distributions in many practical problems, it has serious limitations and inconsistencies. For an understanding of turbulent flow in general, a study of the mean velocity distribution is insufficient. The fields of turbulent fluctuations must be studied in detail. Because turbulent-velocity fluctuations of a fluid are much too complicated, changing too rapidly in time and location to be known in all their details, only a study of some mean values is feasible. These mean quantities include the intensity of turbulent fluctuations, the correlation functions, and the spectrum of turbulence.

Modern statistical theory of turbulence was developed by G. I. Taylor, who introduced the correlation function, spectrum, and the concept of statistically isotropic turbulence. Great simplification may be obtained by the isotropic property. Hence, most of the results from statistical theory of turbulence are concerned with isotropic turbulence.

Correlation function. Consider the fluctuating variables u_A and u_B between stations A and B, and assume that there exists a certain correlation between them. The correlation function is then

$$\rho_{AB} = \overline{u_A u_B} \tag{11}$$

where bar means taking the average. The correlation coefficient R_{AB} is

$$R_{AB} = \frac{\overline{u_A u_B}}{\sqrt{\overline{u_A^2}}\sqrt{\overline{u_B^2}}} \tag{12}$$

The correlation coefficient lies within the limits of -1 and $+1$.

It was von Kármán who first pointed out the tensor character of the correlation function. Both von Kármán and Taylor studied extensively the correlation functions between the components of fluctuating velocity at the same time at two different points of the fluid for isotropic turbulence. These correlation functions had been measured by

hot-wire anemometer. Experimental results check well with theory.

For isotropic turbulence, the correlation function is a function of time t and distance r between two points. The curvature of the double correlation curve at $r = 0$ determines a microscale of turbulence, which is a measure of the size of the smallest eddies in the turbulent flow, these eddies being responsible for the dissipation of turbulent energy. The integration of the correlation coefficient over $0 < r < \infty$ gives the scale of turbulence, which is a measure of the large eddies in the turbulent flow.

Spectrum. A more detailed description of turbulence can be obtained by considering the distribution of energy among eddies of different sizes. This description can be put into precise mathematical form by considering the distribution of energy with frequency or with wave number, which is known as the spectrum of turbulence. Spectral density is a Fourier transform of correlation coefficient. Spectrum of turbulence can be measured by hot-wire anemometer.

Local isotropy. The most significant idea contributed to the problem of turbulent shear flow in recent years is the hypothesis of local isotropy proposed by A. N. Kolmogoroff. He suggested that the fine structure in turbulent shear flow may be isotropic. Turbulent motion is considered to be a mixture of eddies of all sizes from the largest, whose dimensions are comparable with those of the main flow or of the turbulence-producing mechanism such as a grid of bars in a wind tunnel, down to the smallest eddies When turbulent motion starts, the mean flow breaks up, or the eddies produced by the grid break up into smaller eddies, their motions being unstable; these in turn break up into smaller eddies and so on, until eddies are produced of a small enough size to be stable; this gives a lower bound to the eddy size Kolmogoroff's idea may be expressed by saying that there is something universal about small eddies, below a certain eddy size the nature of the motion is unaffected by the origin of the turbulence, and it is expected that eddies, small compared with the dimensions of the mean flow, will be statistically isotropic. Kolmogoroff's idea of local isotropy has been verified experimentally by many research workers

Except for the concept of local isotropy, little has been accomplished for the statistical theory of maintained shear turbulence. However, the statistical theory of isotropic turbulence shows which quantities are important in describing the fluctuating field; they include turbulent intensities correlation function, spectrum, and probability distribution. In the experimental investigations of shear flow, such as flow in circular pipe, channel, boundary layer over a flat plate, jets, and wakes, these are the quantities to be measured.

Turbulent diffusion. Diffusion is a fundamental process of *turbulence There is an essential difference between molecular and turbulent diffusion.* In molecular diffusion, the medium consists of discrete particles, while in turbulent diffusion, the medium is continuous. The old method of investigating turbulent diffusion is semiempirical; it uses Boussinesq's turbulent exchange coefficient. The new method of solution for the turbulent diffusion uses the statistical theory of turbulence. In the statistical theory, two different approaches have been used. One is the continuous stochastic process in which the diffusion equation is obtained from a probabilistic integral equation. The other approach is the random walk method.

Compressible fluid flow. For the turbulent flow of an incompressible fluid, the effect of variation of density in the expression of turbulent stresses is neglected. This effect is no longer negligible for the turbulent flow of a compressible fluid and cannot be neglected for high-speed flow, flow with large variation of temperature, or both The study of the turbulent flow of a compressible fluid requires the correlation of velocity components, of velocity and density, and of pressure and velocity To obtain these three correlations is complicated.

Mixing-length theories may be extended to compressible fluid For two-dimensional parallel flow with mean flow field:

$$\bar{u} = \bar{u}(y) \qquad \bar{v} = 0 \qquad \bar{\rho} = \bar{\rho}(y) \qquad \bar{T} = \bar{T}(y) \tag{13}$$

The fluctuations of velocity component u, density ρ, and temperature T of the fluid may be written as

$$|u'| = l\frac{d\bar{u}}{dy} \qquad |\rho'| = l_\rho \frac{d\bar{\rho}}{dy} \qquad |T'| = l_T \frac{d\bar{T}}{dy} \tag{14}$$

where l, l_ρ, and l_T are the corresponding mixing lengths for the velocity, density, and temperature It is customary to assume that these mixing lengths are equal to simplify the analysis and to aid in solving practical problems. Experimental evidences indicate, however, that they are not equal. For instance, in jet mixing of a compressible fluid, the spread of temperature is wider than that of velocity One way to explain this phenomenon is to assume that l_T is larger than l

The statistical theory of isotropic turbulence has been extended to the case of compressible fluids

Electrically conducting fluid flow. Magnetohydrodynamics deals with flow in electrically conducting fluids in which electromagnetic forces are of the same order of magnitude as gas dynamic forces such as pressure and viscosity Magnetohydrodynamics is important in problems of astrophysics, geophysics, and the behavior of interstellar gas masses, as well as in such engineering problems as reentry of intercontinental ballistic missiles, controlled fusion, and plasma jets Because of the large dimensions, it seems probable that the normal state of motion in the cosmos should be turbulent The high speed of intercontinental ballistic missiles also causes the flow to be turbulent *Controlled fusion research shows* that the turbulent dissipation in magnetohydrodynamics is one of the main difficulties to be overcome before controlled fusion becomes successful In the study

of turbulence in magnetohydrodynamics. correlations between the magnetic field and velocity components play important roles. [S.I.P.]

Bibliography: S. I. Pai, *Viscous Flow Theory*, vol. 2, 1957.

Turkey

Either of two species of North American birds of the family Meleagridae, originally found from Yucatan to Maine. Turkeys are characterized by a lack of feathers on the head and upper neck and by the development of fleshy outgrowths called wattles on the upper neck. The domestic turkey was developed from the Mexican species. The wild

The turkey, *Meleagris gallopavo*, its length is to 50 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

turkey, *Meleagris gallopavo*, differs from the domestic bird in that its tail feathers and upper tail coverts are tipped with chestnut instead of white. It is a bird of the open forest floor, and thrives best where acorns are plentiful. It was a favorite food of the early settlers but was almost exterminated by heavy hunting and habitat destruction. Good progress has been made toward the reestablishment of this colorful American game bird. See GALLIFORMES. [J.D.B.]

Turmeric

This product is both a dye and a spice obtained from the plant *Curcuma longa*, which belongs to the ginger family (Zingiberaceae). It is a stout perennial with short stem, tufted leaves, and short, thick rhizomes which contain the colorful condiment. As a natural dye, turmeric is orange-red or reddish brown, but it changes color in the presence of acids or bases. As a spice, turmeric has a decidedly musky odor and a pungent, bitter taste. It is an important item in curry and is used to flavor and color butter, cheese, pickles, and other articles of food. See SCITAMINALES; SPICE AND FLAVORING. [P.D.S.]

Turmeric plants (*Curcuma longa*) cultivated in Tela, Honduras. (Photo by W. H. Hodge, USDA)

Turn and bank indicator

A device used in aircraft to indicate turning of the aircraft about its vertical axis and the proper bank angle at which to make the turn. The turn indicator is also used to indicate straight flight and the bank indicator to indicate level flight. See AIRCRAFT INSTRUMENTATION.

The turn and bank indicator is composed of two distinct instruments. The bank indicator is essentially a level, consisting of a glass ball rolling in a curved glass tube filled with a damping liquid. It is mounted on the face of the turn indicator. In the absence of lateral accelerations, the deviation of the wings from the horizontal is indicated. Indication off zero during a turn about the vertical axis of the aircraft indicates sideslip.

The turn indicator consists of a rate gyroscope and an angular displacement indicator. The spin axis Y-Y of the gyroscope rotor is parallel to the lateral axis of the aircraft. The gyroscope is permitted to precess against the torque of a restraining spring only about an axis X-X parallel to the principal axis of the airplane. A dashpot for damping is provided. The rotor is driven pneumatically, either by a venturi tube mounted in the slip stream

hot-wire anemometer. Experimental results check well with theory

For isotropic turbulence, the correlation function is a function of time t and distance r between two points. The curvature of the double correlation curve at $r = 0$ determines a microscale of turbulence, which is a measure of the size of the smallest eddies in the turbulent flow, these eddies being responsible for the dissipation of turbulent energy. The integration of the correlation coefficient over $0 < r < \infty$ gives the scale of turbulence, which is a measure of the large eddies in the turbulent flow.

Spectrum. A more detailed description of turbulence can be obtained by considering the distribution of energy among eddies of different sizes. This description can be put into precise mathematical form by considering the distribution of energy with frequency or with wave number, which is known as the spectrum of turbulence. Spectral density is a Fourier transform of correlation coefficient. Spectrum of turbulence can be measured by hot-wire anemometer.

Local isotropy. The most significant idea contributed to the problem of turbulent shear flow in recent years is the hypothesis of local isotropy proposed by A. N. Kolmogoroff. He suggested that the fine structure in turbulent shear flow may be isotropic. Turbulent motion is considered to be a mixture of eddies of all sizes from the largest, whose dimensions are comparable with those of the main flow or of the turbulence-producing mechanism such as a grid of bars in a wind tunnel, down to the smallest eddies When turbulent motion starts, the mean flow breaks up, or the eddies produced by the grid break up into smaller eddies, their motions being unstable; these in turn break up into smaller eddies and so on, until eddies are produced of a small enough size to be stable; this gives a lower bound to the eddy size. Kolmogoroff's idea may be expressed by saying that there is something universal about small eddies, below a certain eddy size the nature of the motion is unaffected by the origin of the turbulence, and it is expected that eddies, small compared with the dimensions of the mean flow, will be statistically isotropic Kolmogoroff's idea of local isotropy has been verified experimentally by many research workers

Except for the concept of local isotropy, little has been accomplished for the statistical theory of maintained shear turbulence However, the statistical theory of isotropic turbulence shows which quantities are important in describing the fluctuating field; they include turbulent intensities, correlation function, spectrum, and probability distribution. In the experimental investigations of shear flow, such as flow in circular pipe, channel, boundary layer over a flat plate, jets, and wakes, these are the quantities to be measured

Turbulent diffusion. Diffusion is a fundamental process of turbulence There is an essential difference between molecular and turbulent diffusion In molecular diffusion, the medium consists of discrete particles, while in turbulent diffusion, the medium is continuous. The old method of investigating turbulent diffusion is semiempirical; it uses Boussinesq's turbulent exchange coefficient The new method of solution for the turbulent diffusion uses the statistical theory of turbulence In the statistical theory, two different approaches have been used. One is the continuous stochastic process in which the diffusion equation is obtained from a probabilistic integral equation. The other approach is the random walk method

Compressible fluid flow. For the turbulent flow of an incompressible fluid, the effect of variation of density in the expression of turbulent stresses is neglected This effect is no longer negligible for the turbulent flow of a compressible fluid and cannot be neglected for high-speed flow, flow with large variation of temperature, or both. The study of the turbulent flow of a compressible fluid requires the correlation of velocity components, of velocity and density, and of pressure and velocity To obtain these three correlations is complicated.

Mixing-length theories may be extended to compressible fluid. For two-dimensional parallel flow with mean flow field:

$$\bar{u} = \bar{u}(y) \qquad \bar{v} = 0 \qquad\qquad \bar{\rho} = \bar{\rho}(y) \qquad \bar{T} = \bar{T}(y) \tag{13}$$

The fluctuations of velocity component u, density ρ and temperature T of the fluid may be written as

$$|u'| = l\frac{d\bar{u}}{dy} \qquad |\rho'| = l_\rho \frac{d\bar{\rho}}{dy} \qquad |T'| = l_T \frac{d\bar{T}}{dy} \tag{14}$$

where l, l_ρ, and l_T are the corresponding mixing lengths for the velocity, density, and temperature It is customary to assume that these mixing lengths are equal to simplify the analysis and to aid in solving practical problems. Experimental evidences indicate, however, that they are not equal For instance in jet mixing of a compressible fluid, the spread of temperature is wider than that of velocity One way to explain this phenomenon is to assume that l_T is larger than l.

The statistical theory of isotropic turbulence has been extended to the case of compressible fluids

Electrically conducting fluid flow. Magnetohydrodynamics deals with flow in electrically conducting fluids in which electromagnetic forces are of the same order of magnitude as gas-dynamic forces such as pressure and viscosity. Magnetohydrodynamics is important in problems of astrophysics, geophysics, and the behavior of interstellar gas masses, as well as in such engineering problems as reentry of intercontinental ballistic missiles, controlled fusion, and plasma jets. Because of the large dimensions, it seems probable that the normal state of motion in the cosmos should be turbulent. The high speed of intercontinental ballistic missiles also causes the flow to be turbulent. Controlled fusion research shows that the turbulent dissipation in magnetohydrodynamics is one of the main difficulties to be overcome before controlled fusion becomes successful In the study

of turbulence in magnetohydrodynamics, correlations between the magnetic field and velocity components play important roles. [S I P.]

Bibliography: S I Pai, *Viscous Flow Theory*, vol. 2, 1957

Turkey

Either of two species of North American birds of the family Meleagridae, originally found from Yucatan to Maine. Turkeys are characterized by a lack of feathers on the head and upper neck and by the development of fleshy outgrowths called wattles on the upper neck. The domestic turkey was developed from the Mexican species. The wild

The turkey, *Meleagris gallopavo*, its length is to 50 in (From E L Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

turkey, *Meleagris gallopavo*, differs from the domestic bird *in that its tail feathers and upper tail* coverts are tipped with chestnut instead of white It is a bird of the open forest floor, and thrives best where acorns are plentiful It was a favorite food of the early settlers but was almost exterminated by heavy hunting and habitat destruction. Good progress has been made toward the reestablishment of this colorful American game bird *See* GALLIFORMES [J.D B]

Turmeric

This product is both a dye and a spice obtained from the plant *Curcuma longa*, which belongs to the ginger family (Zingiberaceae) It is a stout perennial with short stem tufted leaves, and short, thick rhizomes which contain the colorful condiment As a natural dye, turmeric is orange-red or reddish brown, but it changes color in the presence of acids or bases. As a spice, turmeric has a decidedly musky odor and a pungent bitter ...

and ...

of ...

Turmeric plants (*Curcuma longa*) cultivated in Tela, Honduras (Photo by W H Hodge, USDA)

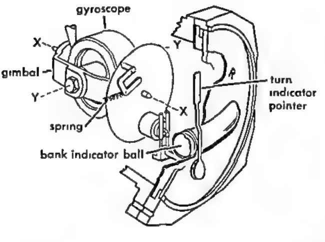
Turn and bank indicator

A device used in aircraft to indicate turning of the aircraft about its vertical axis and the proper bank angle at which to make the turn. The turn indicator is also used to indicate straight flight and the bank indicator to indicate level flight *See* AIRCRAFT INSTRUMENTATION.

The turn and bank indicator is composed of two distinct instruments The bank indicator is essentially a level, consisting of a glass ball rolling in a curved glass tube filled with a damping liquid. It is mounted on the face of the turn indicator ...

... sideslip.

The turn indicator consists of a rate gyroscope and an angular displacement indicator. The spin axis Y-Y of the gyroscope rotor is parallel to the lateral axis of the aircraft. The gyroscope is permitted to precess against the torque of a restraining spring only about an axis X-X parallel ...

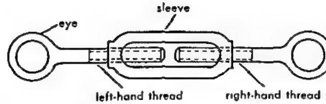


Turn and bank indicator. (*Eclipse-Pioneer Div., Bendix Aviation*)

or by a vacuum pump, or is driven electrically. When the airplane turns about its vertical axis, the rotor precesses an amount approximately proportional to the rate of turn. A pointer attached to the rotor gimbal indicates the angular precession of the rotor. *See* GYROSCOPE. [W.C.B]

Turnbuckle

A device arranged to tighten a rod or wire rope Its parts are a sleeve with a screwed connection at one end and a swivel at the other or, more commonly, with screwed connections of opposite hands (left and right) at each end (as illustrated), so that by



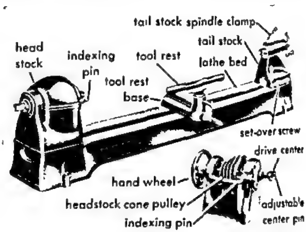
Turnbuckle with eyes.

turning the sleeve the connected parts will be [illegible] nected at any convenient place in the rod or rope and several may be used in series if required [P.H.B]

Turning (woodworking)

The shaping of wood by rotating it in a lathe and cutting it with a chisel. The lathe consists essentially of a bed on which are mounted a headstock, a tailstock, and a tool rest, as illustrated. The headstock is rotated by a motor and holds one end of the wood to be turned The tailstock holds the other end of the wood, allowing it to rotate freely. The tool rest provides a fixed guide along which the operator can handle the chisels if the turning is by hand, or along which the *tool is driven* if the turning is mechanized

Variously shaped chisels are used, all of which have longer handles and thus provide a firmer grip than the woodworking chisels intended to be driven by mallet. A gouge is used for roughing cuts, for



Wood-turning lathe and detail of headstock (*Delta*)

example, in turning the work as it comes from the sawmill to a nearly cylindrical shape. A skew chisel with a straight cutting edge is used in finishing. A parting chisel with a tapered shank is used in separating the finished work from the stock. *See* WOODWORKING. [A T]

Bibliography: H. Hjorth and W. Holtrop, *Operation of Modern Woodworking Machines*, 1958

Turnip

The plant *Brassica rapa* or *B. campestris* var. *rapa* is a cool-season, hardy crucifer of Asiatic origin belonging to the order Papaverales which is grown for its enlarged root or its foliage. The plant is an annual when planted early, a biennial if seeded late in the summer (*see* ANNUAL PLANTS; BIENNIAL PLANTS) Propagation is by seed Popular white-fleshed varieties grown for their roots are Purple Top Globe and White Milan, Yellow Globe and Golden Ball are common yellow-fleshed varieties *See* ROOT (BOTANY), SEED (BOTANY) Shogoin is a popular variety grown principally in the southern United States for turnip greens. *See* LEAF (BOTANY) Turnip harvesting begins when the roots are 2–3 in in diameter, usually 40–70 days after planting. Principal areas of production in the United States are in the South *See* PAPAVERALES; RUTABAGA, VEGETABLE GROWING [H J C]

Diseases of turnip and rutabaga. Turnip and rutabaga are affected by many of the same diseases that attack cauliflower and cabbage (*see* CABBAGE; CAULIFLOWER) Blackleg and black rot are two of the most important. The former infects the roots but causes little damage until the roots are harvested and stored; then a destructive dry rot ensues Similarly, black rot invades the xylem vessels in the roots and is followed by bacterial soft rot in storage. Rotation, clean seed, and hot-water seed treatment will control these diseases

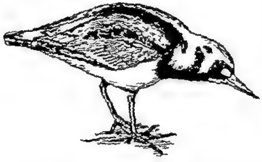
Turnip

Even light infections of clubroot, another disease, are objectionable in turnip and rutabaga because the roots are the parts used for food. After much work resistant varieties of these crops have been produced in Denmark, Sweden, and Great Britain. Boron deficiency is frequently important in turnip and rutabaga causing internal breakdown of the root tissue *See* BORON [C.J.E.]

Bibliography: *see* AGRICULTURAL SCIENCE (PLANT); PLANT DISEASE.

Turnstone

Either of two species of the genus *Arenaria*, of the family *Scolopacidae*. Turnstones differ from other sandpipers in having short legs and rather short, compressed bills, somewhat wedge-shaped and slightly turned upward. The ruddy turnstone, *A. interpres*, is cosmopolitan in distribution, nesting



The ruddy turnstone, *Arenaria interpres*, length about 9 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

in the Arctic. The black ...
phala, ...
winters ...
prefer ...
... being found more commonly on the pebbly beaches, the black on rocky ledges. *See* CHARADRIIFORMES. [J.D.B.]

Turpentine

An essential oil produced by steam distillation of pine woods and from gum turpentine, an exudate of living trees. It is recovered also in the sulfate process of cooking wood pulp. Slash and longleaf pines grown in southeast United States are the chief sources.

Of the essential oils, turpentine is produced in the largest volume. More than 350,000,000 lb is produced annually, of which the United States accounts for about half.

It is a colorless oil whose specific gravity ranges from 0.860 to 0.875. It contains mainly α-pinene, and some β-pinene.

The principal use of turpentine is as a solvent and as a thinner for paints and varnishes. It is a starting material in the production of camphor, borneol, terpineol, and terpin hydrate. *See* CAMPHOR; ESSENTIAL OILS; PINENE. [E.L.S.]

Turquois

A mineral, consisting of hydrated phosphate of aluminum and copper, having the composition $CuAl_6(PO_4)_4(OH)_8 \cdot 4H_2O$ and prized as a semiprecious stone. Fe^{3+} may substitute for some Cu. The name *turquois* means Turkish, and may have been applied because the mineral was first brought to Europe from Persia by way of Turkey. Bone turquois or odontolite, similar to turquois, is formed in fossil bones or teeth and consists of microcrystalline apatite colored by a hydrated phosphate of iron (vivianite). Bone turquois dissolved in hydrochloric acid will not give a blue color when treated with ammonia as will true turquois. The organic origin of bone turquois is readily apparent under a microscope. *See* GEM.

True turquois crystallizes in the triclinic system as short, prismatic crystals, but these are rare. Generally turquois occurs in veinlets or as crusts of massive, dense, finely granular, concretionary, and stalactitic shapes. The color of massive turquois ranges through sky blue, bluish green, apple green, and greenish gray. The robin's-egg blue variety is the most valued. The blue color in turquois is due to the presence of a small amount of copper, the presence of Fe^{2+} results in greenish hues. Some turquois loses color upon exposure to a dry atmosphere. The color may also be altered by treatment with ammonia or acids.

Turquois is a secondary mineral, generally formed in arid regions by the action of surface waters during the alteration of high-aluminous igneous and sedimentary rocks. Phosphoric acid may be derived from the alteration of accessory apatite, and copper may be obtained from disseminated copper sulfides ...

Tu...
local...
... New Mexico, and Nevada. Large deposits located in the Los Cerillos Mountains, about 20 miles southwest of Santa Fe, New Mexico, were mined very early

by Indians and Mexicans and later extensively exploited by Americans. Fine quality turquois has been mined for at least 800 years from a deposit located on the south slopes of the Ali-Mirsa-Kuh Mountains near Nishapur, Iran. Siberia, Turkistan, Asia Minor, the Sinai Peninsula, Silesia and Saxony in Germany, and France possess turquois deposits. [W.R.LO.]

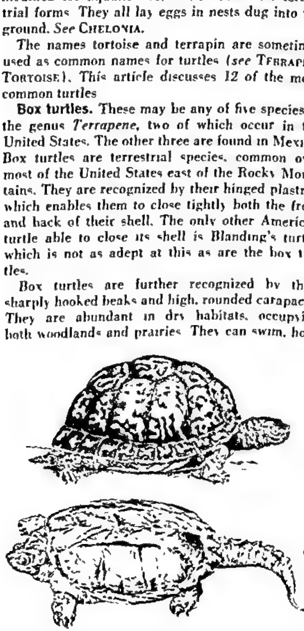
Turtle

Any member of the reptilian order Chelonia (Testudinata) characterized by a dorsal shell, the carapace, and a ventral shell, the plastron, enclosing most of the body. The shell is typically made up of broad, horny plates attached to equally broad, flattened ribs; in some species, the shell is leathery. Instead of teeth, turtles have jaws equipped with sharp, horny plates. Some turtles are highly modified for aquatic life, while others are terrestrial forms. They all lay eggs in nests dug into the ground. *See* CHELONIA.

The names tortoise and terrapin are sometimes used as common names for turtles (*see* TERRAPIN; TORTOISE). This article discusses 12 of the more common turtles.

Box turtles. These may be any of five species of the genus *Terrapene*, two of which occur in the United States. The other three are found in Mexico. Box turtles are terrestrial species, common over most of the United States east of the Rocky Mountains. They are recognized by their hinged plastron which enables them to close tightly both the front and back of their shell. The only other American turtle able to close its shell is Blanding's turtle, which is not as adept at this as are the box turtles.

Box turtles are further recognized by their sharply hooked beaks and high, rounded carapaces. They are abundant in dry habitats, occupying both woodlands and prairies. They can swim, however, and at least one race prefers a damp habitat. Box turtles are omnivorous and exhibit a marked preference for mushrooms. They are also called terrapins, box terrapins, dry-land terrapins, and box tortoises.



Turtles. (a) Box, *Terrapene carolina*; length 5 in. (b) Snapping, *Chelydra serpentina*; length to 3 ft (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Gopher turtles. Any of three species of the genus *Gopherus* are commonly called gopher or gopher tortoise. These turtles have high-arched carapaces flattened dorsally, which are marked with strong concentric ridges. The front legs are covered with enlarged, thickened scales which form a protective shield when the head is withdrawn and the legs are folded into the opening of the carapace. These relatively large, lumbering turtles frequent dry or desert habitats, and dig burrows of some length for protection against extreme heat or cold. Contrary to popular belief, they have a limited home range, seldom wandering more than a few hundred yards from their home burrow. They are vegetarians.

The three species are *Gopherus berlandieri*, the Texas tortoise, found in southern Texas and northeastern Mexico, *G. agassizi*, the desert tortoise ranging in the desert from Nevada and Utah into Mexico; and *G. polyphemus*, the gopher tortoise, a name commonly applied to all of them, found in the southern Coastal Plain from South Carolina into Louisiana and Arkansas.

Green turtles. A large marine animal, known as the green turtle, *Chelonia mydas*, is found throughout the warmer seas of the world. It occurs commonly in the Gulf of Mexico, but also ranges northward in the Atlantic to Massachusetts and in the Pacific to San Diego. Most American specimens weigh from 75 to 150 lb but individuals have been reported to weigh as much as 850 lb.

The green turtle is said to come to land only to lay its eggs, but it also sleeps on isolated beaches. Adults are primarily herbivorous, eating both rooted plants and algae, and occasionally mollusks and crustaceans. The young are more omnivorous than the adults. It is the source of turtle soup and its flesh is highly prized. It is the most important of all turtles commercially.

Loggerhead turtles. Any of several large marine turtles, the loggerhead turtle has limbs modified into flippers, and a large, distinctly beaked head. Best known is the Atlantic loggerhead *Caretta caretta*, found in the Atlantic and Mediterranean which attains a weight of over 500 lb. It is most common along the coast off the mid Atlantic states but ranges northward to Cape Cod in the summer. The eggs and young of this turtle are eaten but adults are too rank for human food. This turtle is carnivorous, eating mollusks and crustaceans and also feeding on the Portuguese man of war.

Map turtles. *These may be any turtle in the genus Graptemys, of which nine species occur in the United States. Some are also called sawbacks. They are moderate-sized aquatic turtles, with a smooth carapace and marked with a distinct median keel. There is considerable range in color of the different species. Best known is the comm[illegible] turtle*

Graptemys geographica. This turtle has a green carapace, sometimes marked with yellow lines and flattened on adults. Mature females have a carapace length of 9–11 in. Males are much smaller. Thin yellow lines on the head and a triangular postorbital mark of yellow are distinguishing marks. It is found throughout the central and upper Mississippi Valley, and eastward into the Lake Ontario region. Related forms occupy most of the area east of the Great Plains. Map turtles exhibit a change in diet with maturity, the young being entirely carnivorous, and most adults preferring a diet of aquatic plants.

Musk turtles. Four similar turtles of the genus *Sternotherus*, all found in the United States, are called musk turtles. They receive their name from their ability to produce a distinctly unpleasant musky odor. Musk turtles are completely aquatic, and only in rare instances do they leave the water except for females who are ready to deposit eggs. They are relatively small, about 3–4 in. in carapace length, with smooth, elongated, highly arched shells, relatively large heads, and a long neck. The plastron is somewhat reduced and does not fill the carapace space. They are eastern and southern in their distribution, only the common musk turtle or stinkpot, *Sternotherus odoratus*, being found north of Arkansas. The latter occurs commonly from Maine to Wisconsin, and southward through Kansas into Texas.

The mud turtles, genus *Kinosternon*, with 14 species, are quite similar. The plastron is somewhat better developed in these than in *Sternotherus*. There are five species in the United States.

Pacific pond turtles. These are sometimes called the western pond turtles; they are the only aquatic turtles with an extensive distribution on the Pacific slope. The species, *Clemmys marmorata*, may be recognized by its olive to brown or black carapace, which is marked with numerous dark, sometimes obscured spots and lines, frequently forming a reticulated pattern on the carapace plates. It reaches a length of 7 in. This turtle is found from Lower California to Oregon and possibly into southern British Columbia. It is a shy species usually found in ponds and quiet pools or streams, but reported in other habitats, including brackish water. It is considered a good food animal, and is frequently sold on the market.

Painted turtles. An abundant aquatic turtle, *Chrysemys picta*, this animal is found throughout the eastern and middle western parts of the U... States ...

Col...

wit...

... carapace is ... green, smooth, and broadly rounded, sometimes marked with yellow. The plastron color varies greatly. The head and neck are variously marked with red, yellow, or orange. The carapace length is usually under 7 in., although the western subspecies is somewhat larger.

Painted turtles are common wherever there is abundant rooted vegetation in quiet, warm, shallow water, including the sluggish portions of weed-filled streams. They are commonly seen sunning on rocks or logs in or near the water. They are omnivorous. This is the turtle most commonly sold in dime stores and pet shops.

Snapping turtles. A large, aquatic animal, the snapping turtle, *Chelydra serpentina*, is common over most of the United States and Canada east of the Rocky Mountains. The snapping turtle is noted for its size and disagreeable disposition. It attains a weight of 60 lb, although the great majority of specimens weigh less than 10 lb. It is dark brown to black above and has a high arching carapace and a large head. The young show three keels on the carapace which become obliterated with age. The snapping turtle is completely omnivorous, and subsists primarily upon crayfish and aquatic vegetation. Although it will eat fish, there is no justification for the popular notion that it is a serious enemy of fishes. Its flesh is of fine flavor, and it is frequently sold on the market.

The alligator snapper, *Macroclemys temmincki*, is similar but much larger, being the largest fresh-water turtle in the United States. This turtle, found in the streams of the southern coastal drainage and well up into the Mississippi Valley, has been reported at weights up to 219 lb. It retains the three keels on the carapace throughout life. There is a related species in Central America and southern Mexico.

Soft-shelled turtles. Any of about 12 species of aquatic turtles of the family Trionychidae are called soft-shelled turtles. They are characterized by a narrow, pointed head and a leathery carapace. The soft-shelled turtles range around the world, occurring in Asia, the East Indies, Africa and North America. There are six distinct types in the United States, usually grouped in two species. One or more types are found in most of the rivers and lakes east of the Rocky Mountains. The back is bronze and the belly white on all of them. They usually have a carapace length of less than 10 in., but much larger specimens occur. These animals have long, flexible necks and a vicious disposition; they will strike quickly and bite fiercely with little provocation.

Soft-shelled turtles feed upon animal food, as well as small amounts of vegetable matter. The flesh is of fine flavor, but these turtles do not live well out of water and are seldom marketed.

Spotted turtles. A semiaquatic species, the spotted turtle, *Clemmys guttata*, is found in bogs, drainage ditches, ponds, and wet woodlands of the eastern United States. It frequently wanders some distance from water, but seems most at home where there is shallow water, mud, and abundant vegetation. This turtle is recognized by its black, smooth rounded carapace, marked with numerous round yellow or orange spots. It is of moderate size, attaining a maximum carapace length of 4½ in. It is primarily insectivorous, but will eat whatever small animal food it can secure. The spotted turtle is common in the northeastern United States, and

is found from New England and southern Ontario, westward across Michigan into Indiana, and south into Georgia.

Wood turtles. The wood turtle, *Clemmys insculpta*, of moderately large size, is recognized by its gray to brown keeled carapace roughened by concentric ridges on the plates. The yellow plastron is marked with a large black blotch on the outer, posterior part of each plate. The average adult has a carapace length of about 7 in. The wood turtle appears to prefer dry woods and meadows during the summer, but in the spring and autumn is more likely to be found leading a semiaquatic life around ponds, streams, and bogs. During long dry spells in the summer, it also tends to seek water. It is entirely omnivorous, but exhibits some preference for vegetable matter, especially fruits, berries, and tender leaves It is found from New England southward into Virginia, and west to central Iowa and Wisconsin. [J D.B.]

Twilight

The period between sunset and darkness in the evening, and between darkness and sunrise in the morning The following statements apply to evening twilight; the reverse would apply to morning twilight.

The characteristic light is caused by atmospheric scattering, which transmits sunlight to the observer for some time after the Sun has set. It depends geometrically on latitude, longitude, and elevation of the observer, and on the time of year. Physically it depends also on local conditions, particularly the weather

Three degrees of twilight are conventionally distinguished Civil twilight ends when the center of the Sun is 6° below the horizon; if the sky is clear it is usually practicable to carry on ordinary outdoor occupations without artificial light during civil twilight. Nautical twilight ends when the depression of the Sun is 12°; at this time both the horizon and the brighter stars are visible Astronomical twilight ends when the depression of the Sun is 18°; at this time no trace of illumination by the Sun appears in the sky. As thus defined, the times of ending of the three sorts of twilight can be precisely calculated. [C M C.]

Bibliography: U.S Naval Observatory, *Tables of Sunrise, Sunset, and Twilight*, supplement to the *American Ephemeris*, 1946.

Twinkling stars

Rapid fluctuation of the brightness and size of the image of a star, caused by the earth's atmosphere, also called astronomical seeing or scintillation Irregularities of temperature and density, accompanied by changes of refractive index, are carried across the beam of light from the star and produce an irregularly corrugated wavefront. Irregularities of the electron density in the ionosphere also change the refractive index at radio frequencies and cause twinkling of extraterrestrial sources of radio noise.

The typical size of the optical irregularities is large compared to the eye but small compared with a telescope. If few are present, or if observed with the eye or a small telescope, they behave like independent lenses or wedges, moving the image bodily and defocusing it without large blurring If many are present, the image is blurred and large Irregularities in low-level air flow contribute to the scintillation, but in good locations the bulk of the seeing is caused by turbulence, which seems to be near the tropopause, at about the 200-millibar pressure level.

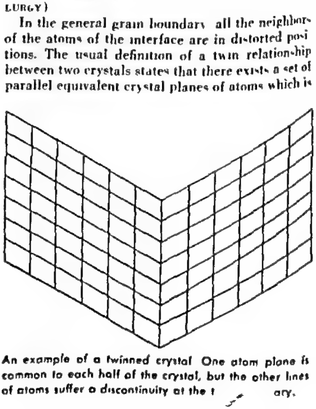
The best astronomical seeing (at night) gives images about 0.″1 of arc; good seeing gives 1″; poor seeing gives up to 10″; daytime seeing is generally worse. Observations up to 40,000-ft altitude indicate that seeing may improve with altitude, and balloon flights have given excellent solar photographs Seeing is worst at large zenith distances, where chromatic effects appear. [J.L.GR]

Bibliography: H. G. Booker and W. E Gordon, *Proc. Inst. Radio Engs.*, 38:401–412, 1950, S Chandrasekhar, *Mon. Not. R.A.S.*, 112:475, 1952, G Keller, *Astrophys. J*, 118:142–146, 1953 (Perkins Obs. Reprint No. 39, 1956).

Twinning (crystallography)

A process in which two or more crystals, or parts of crystals, assume orientations such that one may be brought to coincidence with the other by reflection across a plane or by rotation about an axis. Crystal twins represent a particularly symmetric kind of grain boundary, however, the energy of the twin boundary is much lower than that of the general grain boundary because some of the atoms in the twin interface are in the correct positions relative to each other See GRAIN BOUNDARIES (METALLURGY)

In the general grain boundary all the neighbors of the atoms of the interface are in distorted positions. The usual definition of a twin relationship between two crystals states that there exists a set of parallel equivalent crystal planes of atoms which is



An example of a twinned crystal One atom plane is common to each half of the crystal, but the other lines of atoms suffer a discontinuity at the t ary.

common to both twins, but that rows of atoms are discontinuous across the interface Quite commonly, twins are mirror images of each other, as in the figure Also, it is common for a twin in a crystal to leave the nearest neighbors of the atoms in the interface unchanged in orientation, but to place the atoms in the second neighbor shell in altered positions. This feature is also true of the twin in the figure. It is sometimes possible to create a twin in a crystal by putting an external stress on the crystal, in other cases, twins are found "grown in." See CRYSTAL GROWTH. [R M F.]

Bibliography: R Clark and G B. Craig, Twinning *Progr in Metal Phys*., 3,115–139, 1952.

Tylenchoidea

A superfamily of mainly soil nematodes of small or moderate size with a stylet for piercing live cells and sucking the juices The most important are the phytoparasitic species which attack a wide variety of plants, and often cause extensive damage to agricultural crops Among the most destructive are the stem and bulb eelworms, *Ditylenchus* sp., which attack cereals and flowers, the cyst nematodes *Heterodera* sp., internal root parasites that include the golden nematode, *Heterodera rostochiensis*, a world-wide pest of potatoes, the root-knot nematodes, *Meloidogyne* sp that cause gall formation in many kinds of plants, and the meadow nematodes, *Pratylenchus* sp. that damage roots and expose them to the invasion of bacteria and fungi Most species have fusiform bodies, but in some ectoparasites, such as *Criconema* sp., the body is annulated The gravid females of *Heterodera* and *Meloidogyne* enlarge greatly, the cuticle of the former becomes a globular cyst containing eggs, whereas the females of the latter genus oviposit in a jelly mass There are often races and strains of one species that each parasitize different hosts, but cannot be distinguished morphologically Insect commensals and parasites occur throughout the superfamily, but the Allantonematidae are entirely parasitic in insects Life histories vary greatly, but several stages occur in the insect's hemocoele, and the intestinal and genital systems are damaged by the exit of the nematodes *See* NEMATODA, PLANT DISEASE [H E.W]

Tyndall effect

Visible scattering of light along the path of a beam of light as it passes through a system containing discontinuities The luminous path of the beam of light is called a Tyndall cone In colloidal systems, the brilliance of the Tyndall cone is directly dependent on the magnitude of the difference in refractive index between the particle and the medium In aqueous gold sols, where the difference in refractive index is high strong Tyndall cones are observed

For systems of particles with diameters less than one-twentieth the wavelength of light, the light scattered from a polychromatic beam is predominantly blue in color and is polarized to a degree which depends on the angle between the observer and the incident beam The blue color of tobacco smoke and the blue of the sky are both examples of Tyndall blue As particles are increased in size, the blue color of scattered light disappears and the scattered radiation appears white. If this scattered light is received through a nicol prism which is oriented to extinguish the vertically polarized scattered light, the blue color appears again in increased brilliance. This is called residual blue, and its intensity varies as the inverse eighth power of the wavelength *See* COLLOID; SCATTERING (ELECTROMAGNETIC RADIATION). [Q V.W.]

Type (printing)

Type used in printing is divided into three categories: foundry, machine-cast, and photocomposed type In the first two the face of the letter is raised on one end of a piece of metal. It is from that surface, when inked, that the impression of type is made In photocomposition, the type is reproduced photographically.

Classification. Foundry type, also known as hand type, is cast as single characters in much the same way Johannes Gutenberg, the inventor of movable type, produced them in Mainz, Germany, about 1440

Machine-cast type is produced by Linotype, Intertype, Ludlow, and Monotype machines All but the last cast type in lines, or what are known as slugs The Monotype—in reality two devices, a keyboard and a caster—produces individual types set in lines of desired lengths The Linotype, the

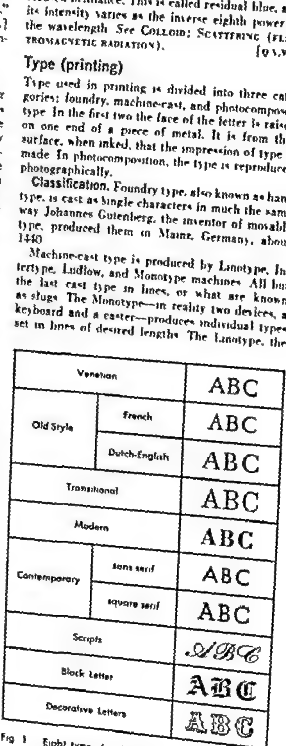


Fig 1 Eight type classifications

is found from New England and southern Ontario, westward across Michigan into Indiana, and south into Georgia.

Wood turtles. The wood turtle, *Clemmys insculpta*, of moderately large size, is recognized by its gray to brown keeled carapace roughened by concentric ridges on the plates. The yellow plastron is marked with a large black blotch on the outer, posterior part of each plate. The average adult has a carapace length of about 7 in. The wood turtle appears to prefer dry woods and meadows during the summer, but in the spring and autumn is more likely to be found leading a semiaquatic life around ponds, streams, and bogs. During long dry spells in the summer, it also tends to seek water. It is entirely omnivorous, but exhibits some preference for vegetable matter, especially fruits, berries, and tender leaves. It is found from New England southward into Virginia, and west to central Iowa and Wisconsin. [J.D.B.]

Twilight

The period between sunset and darkness in the evening, and between darkness and sunrise in the morning. The following statements apply to evening twilight; the reverse would apply to morning twilight.

The characteristic light is caused by atmospheric scattering, which transmits sunlight to the observer for some time after the Sun has set. It depends geometrically on latitude, longitude, and elevation of the observer, and on the time of year. Physically it depends also on local conditions, particularly the weather.

Three degrees of twilight are conventionally distinguished. Civil twilight ends when the center of the Sun is 6° below the horizon; if the sky is clear it is usually practicable to carry on ordinary outdoor occupations without artificial light during civil twilight. Nautical twilight ends when the depression of the Sun is 12°; at this time both the horizon and the brighter stars are visible. Astronomical twilight ends when the depression of the Sun is 18°; at this time no trace of illumination by the Sun appears in the sky. As thus defined, the times of ending of the three sorts of twilight can be precisely calculated. [C.M.C.]

Bibliography: U.S. Naval Observatory, *Tables of Sunrise, Sunset, and Twilight*, supplement to the *American Ephemeris*, 1946.

Twinkling stars

Rapid fluctuation of the brightness and size of the image of a star, caused by the earth's atmosphere, also called astronomical seeing or scintillation. Irregularities of temperature and density, accompanied by changes of refractive index, are carried across the beam of light from the star and produce an irregularly corrugated wavefront. Irregularities of the electron density in the ionosphere also change the refractive index at radio frequencies and cause twinkling of extraterrestrial sources of radio noise.

The typical size of the optical irregularities is large compared to the eye but small compared with a telescope. If few are present, or if observed with the eye or a small telescope, they behave like independent lenses or wedges, moving the image bodily and defocusing it without large blurring. If many are present, the image is blurred and large irregularities in low-level air flow contribute to the scintillation, but in good locations the bulk of the seeing is caused by turbulence, which seems to be near the tropopause, at about the 200-millibar pressure level.

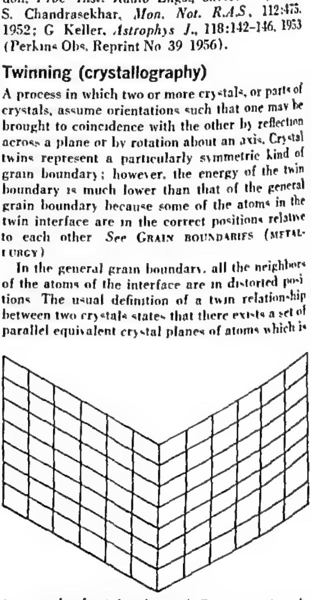
The best astronomical seeing (at night) gives images about 0″.1 of arc; good seeing gives 1″; poor seeing gives up to 10″; daytime seeing is generally worse. Observations up to 40 000-ft altitude indicate that seeing may improve with altitude, and balloon flights have given excellent solar photographs. Seeing is worst at large zenith distances, where chromatic effects appear. [J.L.GR.]

Bibliography: H. G. Booker and W. E. Gordon, *Proc. Inst. Radio Engs.*, 38:401–412, 1950; S. Chandrasekhar, *Mon. Not. R.A.S.*, 112:475, 1952; G. Keller, *Astrophys. J.*, 118:142–146, 1953 (Perkins Obs. Reprint No. 39, 1956).

Twinning (crystallography)

A process in which two or more crystals, or parts of crystals, assume orientations such that one may be brought to coincidence with the other by reflection across a plane or by rotation about an axis. Crystal twins represent a particularly symmetric kind of grain boundary; however, the energy of the twin boundary is much lower than that of the general grain boundary because some of the atoms in the twin interface are in the correct positions relative to each other. See GRAIN BOUNDARIES (METALLURGY).

In the general grain boundary, all the neighbors of the atoms of the interface are in distorted positions. The usual definition of a twin relationship between two crystals states that there exists a set of parallel equivalent crystal planes of atoms which is



An example of a twinned crystal. One atom plane is common to each half of the crystal, but the other lines of atoms suffer a discontinuity at the

Type measurements. Type is cast in sizes from 4 point to 144 point, the sizes in most common use being 6, 7, 8, 9, 10, 11, 12, 14, 18, 24 30, 36, 42, 48, 60, and 72 point.

The American point system, adopted in 1886, made the unit of type measurement a point, 0 01384 in or nearly 1/72 in, and all type sizes are multiples of this unit. Sizes are measured through the body, lengthwise of the letters. This system replaced the sixteenth-century practice of giving all type sizes names such as pearl, agate, nonpareil, brevier, long primer, and pica Some names have remained and have been assigned other functions For example, agate, approximately 5½-point type or 14 lines of type to an inch, has come to be used for measuring advertising space. Publications quote small space rates by the agate line. Nonpareil is used to designate 6 points of space in and around type The word pica is now commonly used to denote a unit of space measuring 12 points It is applied as a dimensional unit to the length of type lines and to the width and depth of pages and blocks of type

The standard height of type in the United States is 0 918 in, a dimension called *type-high* and one observed by photoengravers and electrotypers who make plates to be combined with cast type in let-

... point, are sometimes made of wood

Fonts. A complete complement of letters of one size and style from A to Z, together with the arabic numerals, punctuation, and reference marks, is called a font Most fonts also include the & and $ and, in the body sizes, the five ligatures, ﬁ, ﬂ, ﬃ, ﬀ, ﬄ. One style of letters is used in the setting of body, or ...

tals, itali...

lower cas...

capitals, ... the height of lower-case letters A letter from a different font which is included by accident in type composition is known as a wrong font

Type cases. Foundry type is kept in wooden or metal cases provided with a separate compartment for each letter of a font Before 1900 compositors used two cases, one above the other, the upper case containing the capitals and the lower the small letters, which gave rise to the descriptive expressions, still in use, upper- and lower-case letters

These letters are assembled by hand in a composing stick (small metal tray) and justified (spaced out) with spaces and quads. When the composing stick is full, the contents are transferred to a galley (a flat, oblong open-ended tray) The type matter is held together by several turns of string and is then proved on a proof press After errors have been corrected, it is placed in a chase (metal frame) and held securely by means of

Fig 4 A family of type.

quoins (metal wedges) and furniture and reglets (spacing material lower than type-high), at which point it is ready to be put on the bed of a printing press.

On the Linotype and Intertype, brass matrices with a letter or letters punched in the casting edge are assembled into words and lines by means of a keyboard, and the slug or line of type is cast in one piece from these matrices On the Monotype, coded holes representing the different letters are punched by the keyboard in a 4-in wide roll of paper, and this roll is fed into the caster, where it actuates and controls the casting mechanism.

Type families. A family of type may be likened to the shades of a color in that it includes all variations made of a given type face or design Weights are varied, from light to medium, bold, and extra bold; letters are condensed and expanded, as well as outlined, inlined, and shadowed.

Spacing of type in hand typesetting is done by using thin strips of metal from ½ point and up. When inserted between the lines of type to open them up vertically, it is called leading. When space is inserted between letters of a word to open it up horizontally, it is known as letter spacing.

Special designs. Ornaments and rules (lines) can be cast on typesetting machines. Also found in type metal are individual designs, known as printer's flowers, fleurons, or florets; braces and brackets used to enclose or connect lines of type; chemical, mathematical, astronomical, and medical signs, and signs of the Zodiac.

The arabic numerals introduced into Europe in the twelfth century are made in two ways, old style and modern (or lining). They are part of the fonts of type faces that are derived from designs of the period 1400–1800

Each age since the Italian Renaissance ...

N...

an... Garamond, William Caslon, John Baskerville, Giambattista Bodoni, Frederic W. Goudy, W A. Dwiggins, Rudolf Koch, Paul Renner, Jan van Krimpen, Stanley Morison, and

Defg 72 IJijkl 60

Pmopqr 48 Wstuwxyz 42

Eabcdefghij 36 PQRmnopqrstuv 30

CDEabcdefghijklm 24 FGHIJhijklmnopqrstuv 18

ABCDEFGHIJKLMNOPQuRSabcdefghijklmnopqrstuvwxyz 14

ABCDEFGHIJKLMNOPQuRSTUVabcdefghijklmnopqrstuvwxyz 12

ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz 11

ABCDEFGHIJKLMNOPQuRSTUVWXYZ abcdefghijklmnopqrstuvwxyz 10

ABCDEFGHIJKLMNOPQRSTUVWXYZ abcdefghijklmnopqrstuvwxyz 9

ABCDEFGHIJKLMNOPQuRSTUVWXYZ abcdefghijklmnopqrstuvwxyz 8

ABCDEFGHIJKLMNOPQuRSTUVWXYZ abcdefghijklmnopqrstuvwxyz 6

Fig. 2. Sizes of type

earliest of these composing machines, came into use in 1886 and was invented by Ottmar Mergenthaler.

The product of the photocomposing type machine is the image of type on film or on photosensitized paper in negative and in positive form. This method is called "cold composition" because no hot type metal is needed for its manufacture.

The term cold composition or cold type is also applied to text matter produced on a typewriter or similar machine for reproduction on a printing plate, and to words or lines made up of individual printed characters assembled or pasted together for photographic reproduction.

Of the more than 30,000 type styles created since Gutenberg's time, about 3000 are in everyday use throughout the world. The most widely used method for classifying them is the serif evolution system, based on the different shapes of the terminals or endings of letters. This provides eight classifications: Venetian, Old Style (Dutch-English and French), Transitional, Modern, Contemporary (sans serif and square serifs), Black Letter, Scripts, and Decorative Letters.

ABCDEFGHIJKLMNOPQRSTUVWXYZ&$1234567
890abcdefghijklmnopqrstuvwxyz ff fi ffi fl ffl .:,;-''!?()—
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z & 1234567890
A B C D E F G H I J K L M N O P Q R S T U V
W X Y Z & $ 1 2 3 4 5 6 7 8 9 0 a b c d e f g h i j
k l m n o p q r s t u v w x y z ff fi ffi fl ffl . : , ; - ' ' ! ? ( ) —

Fig 3. A font of type showing roman capitals, roman lower case, small capitals, italic capitals, and italic lower case. Modern and Old Style figures are also shown

Type measurements. Type is cast in sizes from 4 point to 144 point, the sizes in most common use being 6, 7, 8, 9, 10, 11, 12, 14, 18, 24, 30, 36, 42, 48, 60, and 72 point.

The American point system, adopted in 1886, made the unit of type measurement a point, 0.01384 in. or nearly 1/72 in., and all type sizes are multiples of this unit. Sizes are measured through the body, lengthwise of the letters. This system replaced the sixteenth-century practice of giving all type sizes names such as pearl, agate, nonpareil, brevier, long primer, and pica. Some names have remained and have been assigned other functions. For example, agate, approximately 5½-point type or 14 lines of type to an inch, has come to be used for measuring advertising space. Publications quote small space rates by the agate line. Nonpareil is used to designate 6 points of space in and around type. The word pica is now commonly used to denote a unit of space measuring 12 points. It is applied as a dimensional unit to the length of type lines and to the width and depth of pages and blocks of type.

The standard height of type in the United States is 0.918 in., a dimension called type-high and one observed by photoengravers and electrotypers who make plates to be combined with cast type in let-

...ures, punctuation, and reference marks, is called a font. Most fonts also include the & and $ and, in the body sizes, the five ligatures, ff, fi, ffi, fl, ffl. One style of letters is used in the setting of body, or text, matter; it is available in roman capitals, italic capitals (majuscules), roman and italic lower case (minuscules), and an alphabet of small capitals, which is approximately the height of lower-case letters. A letter from a different font which is included by accident in type composition is known as a wrong font.

Type cases. Foundry type is kept in wooden or metal cases provided with a separate compartment for each letter of a font. Before 1900 compositors used two cases, one above the other, the upper case containing the capitals and the lower the small letters, which gave rise to the descriptive expressions, still in use, upper- and lower-case letters.

These letters are assembled by hand in a composing stick (small metal tray) and justified (spaced out) with spaces and quads. When the composing stick is full, the contents are transferred to a galley (a flat, oblong, open-ended tray). The type matter is held together by several turns of string and is then proved on a proof press. After errors have been corrected, it is placed in a chase (metal frame) and held securely by means of

DEFGHI Light
DEFGHI Medium
DEFGHI Bold
DEFGHI Extra Bold
DEFGHI Condensed
DEFGHI Shadowed

Fig. 4. A family of type

quoins (metal wedges) and furniture and reglets (spacing material lower than type-high), at which point it is ready to be put on the bed of a printing press.

On the Linotype and Intertype, brass matrices with a letter or letters punched in the casting edge are assembled into words and lines by means of a keyboard, and the slug or line of type is cast in one piece from these matrices. On the Monotype, coded holes representing the different letters are punched by the keyboard in a 4-in. wide roll of paper, and this roll is fed into the caster, where it actuates and controls the casting mechanism.

Type families. A family of type may be likened to the shades of a color in that it includes all variations made of a given type face or design. Weights are varied, from light to medium, bold, and extra bold; letters are condensed and expanded, as well as outlined, inlined, and shadowed.

Spacing of type in hand typesetting is done by using thin strips of metal from ½ point and up. When inserted between the lines of type to open them up vertically, it is called leading. When space is inserted between letters of a word to open it up horizontally, it is known as letter spacing.

Special designs. Ornaments and rules (lines) can be cast on typesetting machines. Also found in type metal are individual designs, known as printer's flowers, fleurons, or florets; braces and brackets used to enclose or connect lines of type; chemical, mathematical, astronomical, and medical signs; and signs of the Zodiac.

The arabic numerals introduced into Europe in the twelfth century are made in two ways, old style and modern (or lining). They are part of the fonts of type faces that are derived from designs of the period 1400–1800.

Each age since the Italian Renaissance has produced memorable type designers. Among them are Nicolas Jenson, Aldus Manutius (first to use italics and small caps), Claude Garamond, William Caslon, John Baskerville, Giambattista Bodoni, Frederic W. Goudy, W. A. Dwiggins, Rudolf Koch, Paul Renner, Jan van Krimpen, Stanley Morison, and

Bruce Rogers. Many of these men have given their names to the type faces they have designed.

Those who stress type in their design of books and advertising are known variously as typographers, typographic designers, type directors, and type men.

The size of type you are now reading is 9 point and is set 16½ picas wide. *See* COMPOSITION (TYPE); PRINTING PLATE; TYPESETTING. [E.M E.]

Type method

A taxonomic and nomenclatural method for providing a fixed reference point for a taxonomic category. When Linnaeus, the eighteenth century Swedish naturalist, ushered in the modern period of systematic biology through the publication of his classical work, *Systema Naturae*, he established the descriptive method in taxonomy. By means of this method the systematist who recognizes new, or supposedly new, taxa makes these known by means of a technical description and a scientific name. If the new taxon is a species, the material (specimens or parts of specimens) which the author had before him when he described it, together with such additional material as he may have gathered later, was considered typical, and such specimens were known as types Presumably these types represented the author's concept of the species involved and served as standards of comparison and identification At first, the true importance of types was not realized but as knowledge increased and concepts were refined, it was recognized that they were necessary as reference points for the application of scientific names Having a fixed point, later taxonomists could expand or contract or otherwise modify the definition of a given species in the light of subsequent information.

Type designations. For the first 50 or 100 years after Linnaeus, taxonomists did not make formal type designations In fact, Sir James E Smith replaced some of Linnaeus' original specimens with better ones when they became available. Later taxonomists generally designated one specimen to serve as the type (or holotype) for newly described species and selected one, usually called a lectotype, from among competing syntypes or cotypes for the species of earlier authors. As the number of collections and type repositories increased and became less accessible to the taxonomist, as types became lost or destroyed, and as complications arose in trying to apply the single type concept retroactively, more refined type terminology and methods of type designation were developed Although terms are available for more than 100 kinds of "type" specimens, these fall in two major classes (1) nomenclatural types, the name-bearing specimens (onomatophores), which serve as reference points for nomenclatural purposes, and (2) taxonomic types which serve as standards of description and reference points for authors' concepts. The first category has been sanctioned by the International Commission on Zoological Nomenclature, the other by taxonomic practice. Among the more commonly used terms to designate these two kinds of types are the following:

1. Nomenclatural types

a. Holotype: the single specimen designated or indicated as "the type" by the original author at the time of publication of the original description or the only specimen known at the time of the original description.

b Syntype (= cotype): one of several specimens on which an author bases an original description when no single specimen is designated as the holotype.

c. Lectotype: one of a series of syntypes which is selected subsequent to the original description and thenceforth serves as the definitive type of the species. In order to be effective, such selection must be made known through publication.

d. Neotype: a specimen selected as type subsequent to the original description in cases where the primary types are definitely known to be destroyed Here again selections must be made known through publication.

2. Taxonomic types

a. Paratype: a specimen other than the holotype which is before the author at the time of original description and which is designated as such or is clearly indicated as being one of the specimens upon which the original description was based

b Allotype: a paratype of the opposite sex to the holotype which is designated or indicated as such.

c Topotype. a specimen not of the original type series collected at the type locality.

d Plesiotype: a specimen or specimens on which subsequent descriptions or figures are based.

e Metatype a specimen compared with the type by the author of the species and determined by him as conspecific with it

f Homotype: a specimen compared with the type by another than the author of a species and determined by him to be conspecific with it

Stabilization of names. Just as types are required as nomenclatural reference points for the application of scientific names to species, so they are also necessary for the stabilization of names for higher categories This is especially true at the level of the genus, not only because there are more generic names than those of any other higher category but also because generic names provide roots for the names of several categories immediately above them, as tribe, family, and so forth The type for names of a higher category is a lower taxon such as the type of a genus is a species As with species names, the necessity for refining early generic concepts and the retroactive application of the generic type concept to the works of previous authors has resulted in the development of a rather precise series of rules and practices governing their selection. These are specified in detail in the International Rules of Zoological Nomenclature and are concerned with two principal situations. (1) cases in which the generic t ' to be ac-

epted automatically based upon the method or *orm* utilized by the author in the original publication in which the new name was proposed (such as type by original designation, type by monotypy, type by tautonomy), and (2) cases in which a subsequent author has a certain amount of freedom in making a selection from among two or more eligible names for type species (type by subsequent designation). As with other aspects of zoological nomenclature, priority of designation, if properly carried out, is binding upon later authors, except as the International Commission of Zoological Nomenclature, in the interest of conserving current usage, may set aside the rules in a given case. The Commission is also called upon to arbitrate difficult cases of generic type interpretation, as when the author misidentifies the species he designates as the type of a genus and, as a result, the generic diagnosis applies to a species quite different than the one named in his publication. In such cases, the decision is usually influenced by whether the strict application of the author's concept or of a concept based upon the actual species name will have the most adverse nomenclatural effect. *See* ZOOLOGICAL NOMENCLATURE. [E.G.L.]

Bibliography: E. Mayr, E. G. Linsley, and R. L. Usinger, *Methods and Principles of Systematic Zoology*, 1953.

Typewriter

A machine that produces printed copy, character by character, as it is operated. The typewriter may be actuated entirely by hand, in which case the strength of the operator's stroke on a key provides the force to actuate the printing mechanism. The typewriter may contain a powered actuator driven by an electric motor, so that the operator's stroke on a key serves as a control, the actuator providing the force to the printing mechanism. A punched tape may provide the control instead of an operator.

The printed character is produced by a steel type striking an inked ribbon, thereby transferring an impression to paper that rests against a firm platen. An indexing means advances the platen on a carriage after each character is printed, and also advances the ribbon. Usually the ribbon is of fabric, but for sharpest impression, as in preparing copy for photo-offset printing, a ribbon of carbon-coated paper is used. [F.H.R.]

Typhoid fever

A highly infectious, septicemic disease ... [illegible] ... synonym ... typhoid fever is enteric fever. A clinically indistinguishable infection may be caused by other *Salmonellae* (paratyphoid bacilli) and is often referred to as paratyphoid fever (*see* PARATYPHOID FEVER). Typhoid fever must be sharply distinguished from the rickettsial disease typhus (*see* RICKETTSIOSES).

The symptoms of typhoid fever are fever, a rose-colored rash (roseola typhosa), bronchitis, prostration, and a peculiar comatose state. The patient may have either diarrhea or constipation. A majority of the deaths result from complications. The most severe of these are intestinal bleeding, or perforation with peritonitis. Localization of the bacteria in the urinary tract, the gall bladder, bone marrow, and the meninges of the brain may lead to inflammatory complications. The mortality rate varies but is always appreciable.

After recovery, typhoid bacilli may be found in the intestinal tract, the gall bladder, or the urinary tract for a long time. The organisms may thus be excreted in large numbers with stool and urine. Persons so affected are called carriers and are an important source for the spread of infection.

Infection is by the oral route, through the ingestion of food or water contaminated by contact with fecal matter, flies, or unclean hands. The typhoid bacillus can maintain itself in water for long periods of time. Water polluted with fecal matter directly or by leakage from a sewer system is an important source of large epidemics. Milk is second in frequency as a vehicle; other foods follow. Sanitation of water and food are the paramount safeguards against the spread of typhoid fever and other enteric infections. Second in importance are the detection of carriers and their elimination from contact with food.

Where these safeguards are available, incidence of typhoid fever declines close to the vanishing point. For instance, the typhoid death rate per million population in England fell from 388 in 1870 to 0.4 in 1950. Vigilance remains the price of safety. Where sanitation breaks down, as it may during war time or other grave social upheavals, epidemics of enteric infection are certain to reappear.

Diagnosis is based on the clinical symptoms; isolation of the organisms from blood, stools, or urine, and finding of agglutinating antibodies against the causative organisms in blood serum at later stages of the infection or during convalescence (Widal's test).

Infection leaves a considerable degree of immunity. While chloramphenicol is effective in therapy, it unfortunately does not eliminate the carriers (*see* CHLORAMPHENICOL).

A vaccine made of killed typhoid bacilli is widely used for preventive immunization. Broader protection is obtained by TAB vaccine containing, in addition to typhoid bacilli, the paratyphoid organisms most frequently involved in paratyphoid fever. These are *Salmonella paratyphi A* and *B* (in British preparations also *S. paratyphi C*). *See* BACTERIOLOGY, MEDICAL. [A.J.W.]

Typhus, scrub

Scrub typhus, or tsutsugamushi ... [illegible] ... ...erized in man by high

fever, a macular rash, usually a primary eschar, or ulcer, swelling of the regional lymph glands, and development of Weil-Felix OXK agglutinins. *See* RICKETTSIALES; RICKETTSIOSES.

The vectors are larval trombiculid mites The species vary according to location and season, but in most areas are chiefly *Trombicula akamushi* and *T. deliensis* (*see* ACARINA). The primary disease

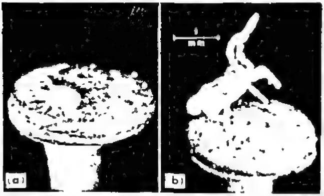


Fig. 1. Trombiculid mites, or chiggers, on pin-heads (a) Engorged, parasitic 6-legged larva feeds only on vertebrates. (b) Velvety, bright red, 8-legged adult is not parasitic and not a transmitter of disease except indirectly through larval progeny. (From C B Philip, Scrub typhus, Sci. Monthly, 69 287, 1949)

cycle is between these mites, which are parasitic on vertebrates only in the larval stage, and field mice and rats. Transmission to man occurs from the next generation of larval mites.

As with spotted fever, scrub typhus varies in virulence in different foci with fatality rates ranging from below 10 to over 60%. In the north, the disease is seasonal, and new foci have been discovered in Japan and Korea, but in the tropics and subtropics from India as far east as northeastern Australia and the Solomon Islands, incidence is by exposure more than by season.

Strains of *R. tsutsugamushi*, even those in a small local area, do not have the same antigen complement. This is referred to as antigenic heterogeneity and is considered one of the factors in recurrent attacks of the disease. Two strains have been reported as elaborating a toxin. *See* ANTIGEN; TOXIN, BACTERIAL.

The agent grows rather poorly in injected arthropods Although *R tsutsugamushi* grows well in the yolk sacs of embryonated eggs and in lungs of white mice and cotton rats, vaccines produced in these hosts have proven unsatisfactory because of the great antigenic heterogeneity displayed by this organism The vaccines so produced would not be effective against an infection due to a strain with a

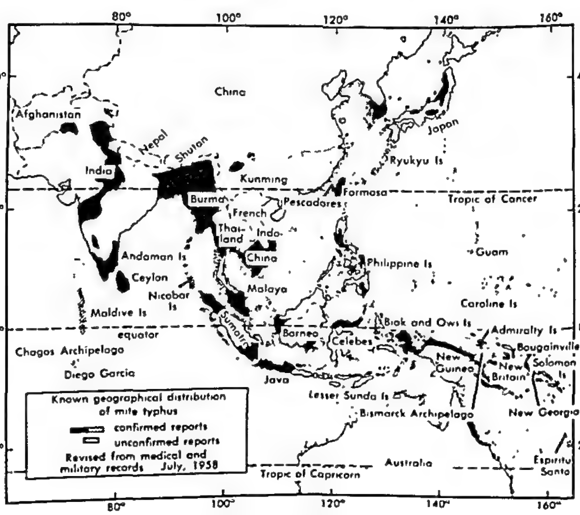


Fig 2 The endemic occurrence of scrub typhus in the Far East and South Pacific regions

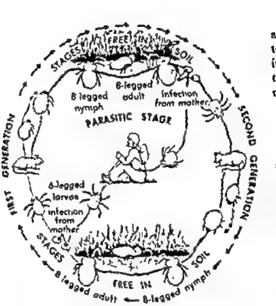


Fig. 3. Two-year natural cycle of scrub typhus. (*From C. B. Philip, Scrub typhus, Sci. Monthly, 69:289, 1949*)

different antigenic complement. However, following demonstration of efficacy of antibiotic therapy, immunization has been accomplished by inducing infections in man and then treating the infection with antibiotics. [C.B.P.]

Typhus fever, endemic (flea-borne)

An infectious disease closely related to epidemic typhus fever. At one time, endemic typhus was thought to be a variant of the same infection in the New World where it was first differentiated, but it is now accepted as specifically distinct and also world wide. This generally milder, febrile disease with more gradual onset, shorter duration, and lower mortality than the epidemic form is caused by *Rickettsia typhi* (synonyms *manchuriae*, *mooseri*). See RICKETTSIALES; RICKETTSIOSES; TYPHUS FEVER, EPIDEMIC (LOUSE-BORNE).

The primary disease cycle occurs between rats and their fleas with sporadic involvement of man. There is much wider host susceptibility range than with *R. prowazekii*. In contrast to *R. prowazekii*, *R. typhi* produces a characteristic scrotal ... in gu... nonre... persi... experimentally in brains of white rats and ...

... these two. The complement-fixing and agglutinating antibodies, as well as toxic properties, are specific. See SEROLOGY.

The vector is chiefly the tropical rat flea, *Xenopsylla cheopis*, though other fleas can be infected (see SIPHONAPTERA). As in epidemic typhus, contamination by the feces of the vector, rather than its bites, is responsible for transmission.

Because of the primary cycle in commensal rats and their fleas, prevention and control are directed to them. The sporadic human incidence does not indicate need for a vaccination program, and protection of personnel in food-handling establishments in urban and suburban areas, where most ... by rat control ...

... areas, but whether these measures are responsible for decreasing incidence in southeastern United States cannot be assessed critically. For example, there has been a concomitant decrease in Puerto Rico where such measures have not been in progress. [C.B.P.]

Typhus fever, epidemic (louse-borne)

An acute, febrile human disease caused by bacteria-like microorganisms, *Rickettsia prowazekii*, and transmitted by the human louse, *Pediculus humanus*. It is the most universally known and longest recognized of the rickettsioses because of its past ravages during wars and famine. History abounds in spectacular accounts of its adverse effects on major military campaigns, such as during the Napoleonic wars and in Serbia at the beginning of World War I. See RICKETTSIALES; RICKETTSIOSES.

Its harborage typically is among distressed, lousy populations. The lower animals are not a part of the reservoir mechanism, as in endemic typhus, though monkeys, guinea pigs, and African gerbilles have been infected, and white mice, white rats, and rabbits sustain so-called inapparent infections. See TYPHUS FEVER, ENDEMIC (FLEA-BORNE).

Clinical picture. In man, some 10–14 days after exposure, onset is usually abrupt with headache, malaise, and generalized aches and pains. Fluctuating but sustained fever between 100–105°F lasts over 2 weeks. A generalized macular rash occurs

Rickettsia prowazekii of epidemic typhus in infected yolk cell of chicken embryo showing pleomorphism (filaments, short rods, and coccoid forms). (*Photomicrograph by N. J. Kramis*)

about the fourth or fifth day, accompanying marked prostration. The second to third weeks of illness are critical, the patients frequently becoming stuporous and disoriented. Fatality rates vary but average about 20% and have reached as high as 70% in severe epidemics. Recovery is surprisingly rapid and sustained after-effects are rare for so severe an illness accompanied by lesions in the brain. Immunity is generally lasting, but recrudescences many years afterward in the absence of lice are known to occur, as Brill's disease.

Etiologic agent. *The etiologic agent, R. prowazekii, was named for two early investigators who lost their lives from this infection. The organisms are minute coccobacillary or rod-shaped cells which may form filaments up to 40 μ in length. In general, they vary from 0.3–0.7 μ in diameter by 0.5–2.0 μ long. They grow abundantly within cells lining the midgut of the louse host and the blood vessels of man. Their vigorous proliferation in yolk sacs of embryonated chicken eggs has provided the basis for the present, widely used Cox-type vaccine.*

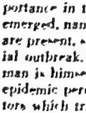
Insect vector. *The insect vector, P. humanus, has two varieties, the body louse, var. humanus, sometimes called corporis or vestimenti, and the head louse, var. capitis, both of which become infected although the former is the more important (see ANOPLURA). The infection is lethal to the louse in about 9 days. One outstanding achievement of preventive medicine during World War II was the demonstration of effectiveness of DDT insecticide against body lice during the Naples typhus epidemic in late 1943. However, this preventive measure later became complicated by discovery that strains of lice in Korea and elsewhere had become insecticide-resistant. This placed greater emphasis on vaccination for military requirements. Source of infection is not by louse bite but by contamination of the bite wound, or less often from bedding and clothing, by infected louse feces.*

Diagnosis. *Diagnosis is assisted by history of lousiness, evolution of the rash from the trunk to the extremities, and supplemented by laboratory procedures. The agent can be recovered during fever by injection of blood into chick embryo yolk sacs or guinea pigs (which develop fever but no scrotal involvement), and by feeding clean lice on the patient. The Weil-Felix OX_{19} antibodies appear in about a week, rise to a maximum about 2 weeks later, and usually disappear in a few more weeks. Complement-fixing antibodies appear about the same time, rise to a peak after 2–3 weeks' convalescence, and decrease slowly over a period of some months; specific agglutinins show much the same activity. Specific toxin neutralizing antibodies can also be demonstrated. See ANTIBODY; SEROLOGY; TOXIN, BACTERIAL.*

Brill's disease. *Brill's disease is now accepted as a mild, recurrent form of epidemic typhus in persons previously infected, sometimes many years before. It is named for the physician who first reported these moderated, typhuslike cases in New York City in the absence of lice and with never more than one case to a family. H. Zinsser later observed that 90% of U.S. cases were in European emigrants; thirteen of his cases had actually emigrated from Central Europe over 30 years previous to their later episodes of illness. Recent studies in the United States and Europe, particularly by Murray and associates in Yugoslavia, have abundantly confirmed Zinsser's hypothesis, and a fact of importance in the puzzling reservoir mechanism has emerged, namely, that in a household in which lice are present, such a case can occur prior to a familial outbreak. This has led to the conclusion that man is himself supplying the source for the interepidemic persistence of this agent, though the factors which trigger the recurrence remain obscure.*

[C.E.F.]

Tyrosine



Tyrosine

Physical constants of the L isomer at 25°C

pK_1 (COOH) 2.20; pK_2 (NH_3^+) 9.11

Isoelectric point 5.66

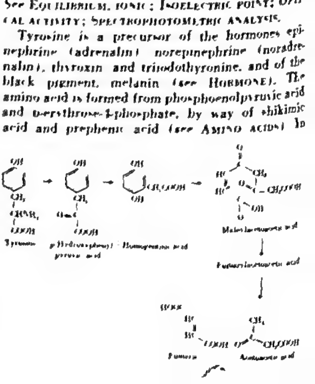
Optical rotation $[\alpha]_D$ (5 N HCl) −10.0

Solubility (g/100 ml H_2O) 0.05

Absorption spectrum peak at 275 mμ (ultraviolet)

An amino acid. The amino acids are characterized physically by the following: (1) the pK_1, or the dissociation constant of the various titratable groups; (2) the isoelectric point, or pH at which a dipolar ion does not migrate in an electric field; (3) the optical rotation, or the rotation imparted to a beam of plane-polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml; (4) solubility; (5) absorption spectrum, or the wavelength at which maximum absorption occurs. See EQUILIBRIUM, IONIC; ISOELECTRIC POINT; OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS.

Tyrosine is a precursor of the hormones epinephrine (adrenalin), norepinephrine (noradrenalin), thyroxin and triiodothyronine, and of the black pigment, melanin (see HORMONE). The amino acid is formed from phosphoenolpyruvic acid and D-erythrose-4-phosphate, by way of shikimic acid and prephenic acid (see AMINO ACIDS). In



animals, tyrosine is formed by the oxidation of dietary phenylalanine. The major pathway for metabolic degradation leads to fumaric and acetoacetic acids, by the reactions shown. *See* PHENYLALANINE. [E.A.AD.]

Tyrothricin

The first antibiotic to be used for human therapy. It was isolated from a culture of *Bacillus brevis* by Rene J. Dubos in 1939. This basic discovery laid the groundwork for the many antibiotics which have followed (*see* ANTIBIOTIC).

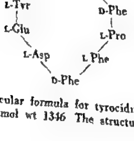
Tyrothricin is a combination of two separate antibiotics. At least 18% is gramicidin, and 65% is tyrocidine hydrochloride. The remainder consists of other polypeptides. Crystalline gramicidin, a cyclic polypeptide with a melt...

formula is uncertain because it is not attacked by any known enzyme...

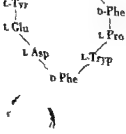
mula is $C_{144}H_{210}O_{26}N_{10}$.

Another compound, Gramicidin S, resembles closely the known tyrocidines in both clinical and microbiological properties. Despite the name, however, Gramicidin S is not a component of tyrothricin as described by Dubos. It is derived from a species closely related to *Bacillus brevis*, the organism which makes tyrothricin.

Crystalline tyrocidine is a mixture of polypeptides containing at least three components A, B, and C. The molecular and structural formulas for components A and B have been determined. The molecular formula for tyrocidine A is $C_{66}H_{86}O_{13}N_{13}$, molecular weight (mol wt) 1270. The structural formula follows



The molecular formula for tyrocidine B is $C_{68}H_{88}O_{13}N_{14}$, mol wt 1346. The structural formula follows

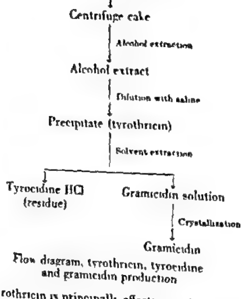


In the formulas Phe = phenylolanine, Asp = aspartic acid, Glu = glutamic acid, Leu = leucine, Orn = ornithine, Pro = proline, Tryp = tryptophan, Tyr = tyrosine, and Val = valene.

Tyrothricin is a white, amorphous powder which is insoluble in water, chloroform and ether, but soluble in many organic solvents, such as alcohol and propylene glycol. Clear, stable aqueous suspensions may be prepared by adding minute quantities of surface active agents. The outstanding thermostability of tyrothricin permits sterilization at normal autoclave temperatures without any apparent loss in activity.

United States production of tyrothricin in 1956 amounted to...

cells of *Bacillus brevis* with alcohol. Dilution of this extract with saline precipitates tyrothricin.



Flow diagram, tyrothricin, tyrocidine and gramicidin production

Tyrothricin is principally effective against a wide variety of gram-positive organisms. In test tube studies (in vitro), tyrothricin is effective in concentrations as low as 1 microgram per milliliter ($\mu g/ml$). At higher concentrations, 150–500 $\mu g/ml$, many gram negative and fungal organisms are susceptible in vitro. Commercial products usually contain 200 $\mu g/ml$ for nasal preparations, 1000 $\mu g/g$ for topical preparations, and 2000 $\mu g/g$ for throat lozenges.

In the living organism, tyrothricin has an extremely low incidence of side effects. Sensitization by tyrothricin is exceedingly rare. The safety of tyrothricin is emphasized by the fact that batch certification and expiration dating are unnecessary. Tyrothricin products may be sold over the counter without a physician's prescription.

Since tyrothricin introduced into the blood stream causes hemolysis, use of this antibiotic has been limited to topical preparations. Investigators

have reported that tyrothricin assists markedly in the formation of granulation tissue and the promotion of wound healing.

Tyrothricin is offered commercially in ointments, creams, powders, lotions, sprays, suppositories, tablets, and aerosol preparations Such products are used for the treatment of dermal infections, infections of the nose, throat and eyes, and for the preparation of solutions used to irrigate the body cavities. [B C M]

Bibliography: R. D. Hotchkiss, Gramicidin, tyrocidine, and tyrothricin, *Advances in Enzymol*, 4:153–199, 1944.

U Ulcer to Uterus

Ulcer

Common inflammatory lesions in which there is a loss or destruction of superficial tissue. Ulcers occur in several locations as acute, subacute, chronic, or recurrent types. In each location some initiating factor, such as action of bacterial toxins or lack of oxygen, causes death of the surface tissue. This necrotic tissue then usually sloughs off, leaving the underlying area exposed to further damage. Some inflammation is present. See TOXIN, BACTERIAL.

Peptic ulcers are a common form of ulcer which occur primarily in the stomach and duodenum. The exact causes are obscure, although predisposing factors such as tension, anxiety, and abnormal amounts of acid secretion appear in most cases. Most of these tend to be chronic or recurrent in nature and the clinical course is quite variable, depending upon the individual. Although the fatality rate is low, serious consequences may result from perforation, hemorrhage, or obstruction. In addition, carcinoma appears at the ulcer site in a very small percentage of patients.

A special form of gastric ulcer is seen after severe body burns and is called Curling's ulcer; similar lesions appear in other cases when severe damage or stress are incurred.

Ulcerative colitis of the large intestine is similar, in many respects, to peptic ulcer, because its exact cause is unknown and a relationship exists with certain personality patterns.

In many skin disorders in which a circulatory deficiency is involved, ulcerative lesions may appear. These also tend to be chronic and are seen most often in patients with arteriosclerosis, diabetes mellitus, and varicosities, and as a result of physical injury, such as frostbite and trench foot. In each, a tissue ischemia is produced, the lower limbs are most often involved.

Superficial ulcerations of the mouth, tongue, and other mucosal surfaces may occur as the result of illness, infection, or irritation. Several serious diseases may be accompanied by characteristic ulcers of the gastrointestinal tract, including typhoid fever, tuberculosis, and both amebic and bacillary dysentery. See AMEBIASIS, BACILLARY DYSENTERY; TUBERCULOSIS, TYPHOID FEVER. [E.C.S.T.]

Ulmin

A term used to designate alkali-soluble organic substances derived from decaying vegetable matter. The term has been replaced by humus and humic substances with which ulmin is synonymous. See HUMUS; see also COAL, PEAT. [I.A.B.]

Ultimobranchial bodies

Small glandular structures which originate as terminal outpocketings from each side of the embryonic pharynx. They occur only in vertebrates, where they are almost universal but difficult to homologize. They probably represent an expression of continued growth activity, associated with pouch- or gill-forming potentialities of the foregut entoderm. They are usually bilateral in mammals. The last, and sometimes next to last, "true" pouch and ultimobranchial primordium often unite to form evaginations, known as the caudal pharyngeal complex. In man, they may be related to the third, as well as the fourth, pharyngeal pouch.

In fishes and amphibians, these structures have been known as suprapericardial bodies because of their position. Other terms, employed to designate these structures, are the postbranchial and telobranchial bodies. Originally regarded as vestigial lateral, or accessory, thyroids, because in mammals the ultimobranchial bodies were found to join with lateral lobes of the thyroid, this tissue has more recently been interpreted as relatively indifferent and capable of modification by various factors. When intimately incorporated within a growing, mammalian thyroid or sometimes even when outside that gland, as in the baboon, it is usually indistinguishably transformed into thyroidlike tissue, but appears to function as such only during periods of thyroid activity. This may occur experimentally under the influence of thyrotrophin.

Structure. Morphologically, ultimobranchial tissue can be highly variable. In fishes and amphibians, it is often vesicular, but evidence for secretory activity occurs mostly in reptiles and birds. In mammals, it may reflect features more typical of lower forms. Within old, or atrophic, thyroids, it is sometimes cystic and nonsecretory. It can become mucus-secreting, which is reminiscent of its entodermal origin.

Function and fate. Ultimobranchial bodies, at least in mammals, appear to possess few, if any, specific functions. The proximity of these bodies to true pouches during development suggests they can carry induced, if not intrinsic, attributes. Since foregut entoderm may have reciprocal potentialities, this tissue may produce thymus or parathyroid tissue.

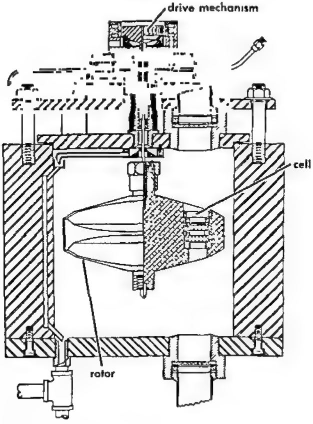
This tissue can be experimentally induced to transform into multiple cysts composed of stratified squamous epithelium, a process called metaplasia. Such metaplasia, however, is often reversible. In human beings, ultimobranchial tissue may become

tumorous under the influence of factors causing thyroid tumors. In rats such tumors can be produced experimentally. *See* EMBRYONIC INDUCTION; METAPLASIA. [J.H.V.D.]

Bibliography: J. H. Van Dyke, Behavior of ultimobranchial tissue in the postnatal thyroid gland, *Am. J. Anat.*, 76(2):201–243, 1945; J. H. Van Dyke, Experimental thyroid metaplasia in the rat, *A.M.A. Arch. Pathol.*, 59:73–81, 1955.

Ultracentrifuge

A centrifuge of low or high speed which provides convection-free conditions and a quantitative means for measurement of sedimentation velocity, sedimentation equilibrium of solutes, or both, in liquid solutions. The term supercentrifuge is used for centrifuges of high speed which fail to meet either of these two criteria



An air-driven ultracentrifuge.

The ultracentrifuge is used to measure molecular weights of solutes and to provide data on particle size or molecular-weight distributions in polydisperse systems

The rotor, cell, drive mechanism, and temperature control are designed to achieve conditions of convectionless sedimentation and high centrifugal accelerations. Optical systems have been designed for recording concentration c versus distance x in the centrifuge cell and concentration gradient dc/dx versus distance x in the cell while the centrifuge is in operation Four methods for molecular-weight measurements are available

Sedimentation velocity method. This method utilizes centrifugal accelerations which are great enough to produce conveniently measurable sedimentation rates for solutes. A sedimentation constant s is obtained which is defined by the equation

$$s = \frac{dx/dt}{\omega^2 x}$$

where dx/dt is sedimentation velocity (cm/sec), ω is angular velocity of the rotor (rad/sec), and x is distance from center of rotation (cm) of a reference point in the sedimenting system Molecular weight M is obtained from the relationship

$$M = \frac{RTs}{D(1 - V\rho)}$$

where R is the gas constant, T is the absolute temperature, D is the diffusion coefficient of the solute V is the partial specific volume of the solute, and ρ is the density of the solvent.

Sedimentation equilibrium. When an ultracentrifuge is operated at speeds too low to cause measurable sedimentation rates, but sufficiently high to produce a redistribution of solute in the centrifuge cell, a condition of sedimentation equilibrium may be established after a period of time. The relationship between concentration distribution and molecular weight at equilibrium is

$$M = \frac{2RT \ln c_2/c_1}{(1 - V\rho)\omega^2(x_2^2 - x_1^2)}$$

where c_1 and c_2 are concentrations of solute at levels x_1 and x_2 For high-molecular-weight solutes, times up to several weeks of continuously running the ultracentrifuge may be required to establish equilibrium

Archibald approach-to-equilibrium method. The relations governing concentration distribution at equilibrium govern the concentration distribution at the ends of the cell column in the ultracentrifuge throughout the approach to equilibrium. Therefore, an equilibrium result for molecular weight may be obtained from the concentration distribution at either end of the solution column long before complete equilibrium is reached

Vinograd density-gradient method. Binary solutions of low-molecular-weight components of different densities will exhibit a gradation in density with distance x in the cell column at sedimentation equilibrium. When a macromolecular species is present in such a density gradient, it will take up a position in the cell where its effective density is matched by the density of the medium. The width of the band of macromolecules at the equilibrium density position is related to the molecular weight of the macromolecule *See* CENTRIFUGATION; COLLOID; MOLECULAR WEIGHT. [Q.V.W.]

Ultrafiltration

A filtration process in which particles of colloidal size are retained by a filter medium while solvent

plus accompanying low-molecular-weight solutes are allowed to pass through. Ultrafilters are used (1) to separate colloid from suspending medium, (2) to separate particles of one size from particles of another size, and (3) to determine the distribution of particle sizes in colloidal systems by the use of filters of graded pore size.

Ultrafilter membranes have been prepared from various types of gel-forming substances. Unglazed porcelain has been impregnated with gels such as gelatin or silicic acid. Filter paper has been impregnated with glacial acetic acid collodions of varying strengths to produce a series of filters of graded porosity.

A new type ultrafilter membrane is made up of a thin plastic sheet containing millions of tiny pores evenly distributed over its surface. Flow rates of liquids and gases through these membranes are very high because the pore volume is 80% of the total membrane volume and the pores proceed through the filter in a direct path. Nominal pore diameters may be obtained in the range from 10 mμ to 5.0 μ. *See* COLLOID; FILTRATION [Q.V.W.]

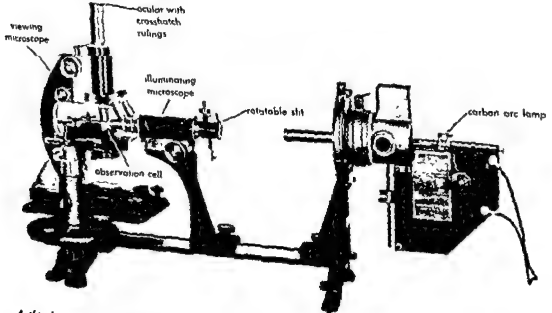
Ultramicroscopy

Investigation of particles of submicroscopic dimensions with an ultramicroscope. An ultramicroscope (see illustration) consists of a high-intensity illumination system for producing a Tyndall cone in a [illegible] *scattered by* the particles enters the microscope (dark field illumination). The particles are too small to be resolved by the microscope, but the points of light emanating from the particles reveal their number, position and Brownian motion. The factors which determine the lower size limit of particles which may be detected in an ultramicroscope are the intensity of the illumination and the difference in refractive index between particle and medium. The magnification provided by the viewing microscope is of secondary importance. High-intensity illumination is obtained either from the sun by means of a heliostat or from a clock-feed carbon arc lamp.

The field area within which particle counts are made is defined by the crosshatched rulings in the ocular. The depth of illumination is controlled by the adjustable slit, and is measured by rotating the slit through 90° and observing the width of the Tyndall cone with the ruled ocular. The product of field area by depth of illumination gives the size of the volume element which corresponds to a particle count.

H. Siedentopf and R. Zsigmondy were able to count particles of gold in water as small as 8 millimicrons (mμ) in diameter using a carbon arc and as small as 5 mμ using the sun's rays. They determined the average particle mass in colloidal systems by counting the number of particles in a known volume and, in a separate experiment, determining the total mass of particles in a known volume. The ratio, mass/number, gave the average particle mass. This result was converted to a particle dimension by making assumptions for particle density and shape.

In many colloidal systems, the difference in refractive index between particle and medium is not sufficiently great to produce visible scattering from individual particles in the ultramicroscope. The ultramicroscope has been used principally in studies of metallic particles in various liquid and transparent solid mediums. *See* BROWNIAN MOVEMENT; COLLOID; MICROSCOPE; TYNDALL EFFECT. [Q.V.W.]



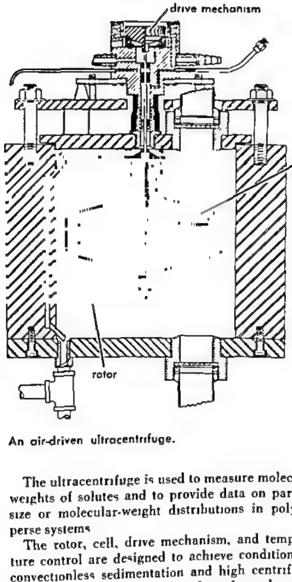
A slit ultramicroscope (Bausch and Lomb)

tumorous under the influence of factors causing thyroid tumors. In rats such tumors can be produced experimentally. *See* EMBRYONIC INDUCTION; METAPLASIA. [J.H.V.D.]

Bibliography: J. H. Van Dyke, Behavior of ultimobranchial tissue in the postnatal thyroid gland, *Am. J. Anat.*, 76(2):201–243, 1945; J. H. Van Dyke, Experimental thyroid metaplasia in the rat, *A.M.A. Arch. Pathol.*, 59:73–81, 1955.

Ultracentrifuge

A centrifuge of low or high speed which provides convection-free conditions and a quantitative means for measurement of sedimentation velocity, sedimentation equilibrium of solutes, or both, in liquid solutions. The term supercentrifuge is used for centrifuges of high speed which fail to meet either of these two criteria.



An air-driven ultracentrifuge.

The ultracentrifuge is used to measure molecular weights of solutes and to provide data on particle size or molecular-weight distributions in polydisperse systems.

The rotor, cell, drive mechanism, and temperature control are designed to achieve conditions of convectionless sedimentation and high centrifugal accelerations. Optical systems have been designed for recording concentration c versus distance x in the centrifuge cell and concentration gradient dc/dx versus distance x in the cell while the centrifuge is in operation. Four methods for molecular-weight measurements are available.

Sedimentation velocity method. This method utilizes centrifugal accelerations which are great enough to produce conveniently measurable sedimentation rates for solutes. A sedimentation constant s is obtained which is defined by the equation

$$s = \frac{dx/dt}{\omega^2 x}$$

where dx/dt is sedimentation velocity (cm/sec), ω is angular velocity of the rotor (rad/sec), and x is distance from center of rotation (cm) of a reference point in the sedimenting system. Molecular weight M is obtained from the relationship

$$M = \frac{RTs}{D(1 - V\rho)}$$

where R is the gas constant, T is the absolute temperature, D is the diffusion coefficient of the solute, V is the partial specific volume of the solute, and ρ is the density of the solvent.

Sedimentation equilibrium. When an ultracentrifuge is operated at speeds too low to cause measurable sedimentation rates, but sufficiently high to produce a redistribution of solute in the centrifuge cell, a condition of sedimentation equilibrium may be established after a period of time. The relationship between concentration distribution and molecular weight at equilibrium is

$$M = \frac{2RT \ln c_2/c_1}{(1 - V\rho)\omega^2(x_2^2 - x_1^2)}$$

where c_1 and c_2 are concentrations of solute at levels x_1 and x_2. For high-molecular-weight solutes times up to several weeks of continuously running the ultracentrifuge may be required to establish equilibrium.

Archibald approach-to-equilibrium method. The relations governing concentration distribution at equilibrium govern the concentration distribution at the ends of the cell column in the ultracentrifuge throughout the approach to equilibrium. Therefore an equilibrium result for molecular weight may be obtained from the concentration distribution at either end of the solution column long before complete equilibrium is reached.

Vinograd density-gradient method. Binary solutions of low-molecular-weight components of different densities will exhibit a gradation in density with distance x in the cell column at sedimentation equilibrium. When a macromolecular species is present in such a density gradient, it will take up a position in the cell where its effective density is matched by the density of the medium. The width of the band of macromolecules at the equilibrium density position is related to the molecular weight of the macromolecule. *See* CENTRIFUGATION; COLLOID; MOLECULAR WEIGHT. [Q.V.W.]

Ultrafiltration

A filtration process in which particles of colloidal size are retained by a filter medium while solvent

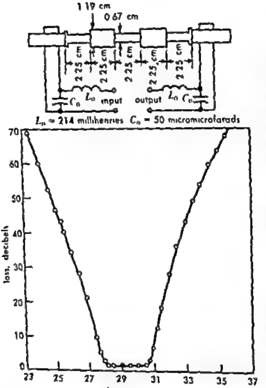


Fig 2 Torsional wave electromechanical wave filter and resulting attenuation

(cylinders of metal with lengths and diameters as shown) connected to and driven by two *torsional* transducers.

Ultrasonic inspectoscopes These transmit sound waves into a metal casting or other solid piece and determine the presence of flaws by reflections or by an interruption of the sound-wave transmission through the piece Frequencies ranging from 500 kc to 15 Mc are used Such devices are among the best means for determining defects in metals, glasses, and ceramics, and they have also been applied in the inspection of automobile tires See METAL INSPECTION, ULTRASONIC

Ultrasonic thickness gages These have been used in measuring the thickness of pieces for which one side

syste

on th

Hi

acou waves have been used in a variety of applications involving gases, liquids, and solids Some of the more common applications are mentioned here

Effects due to cavitation. Holes (gas-bubble cavities) can be created in a liquid by high-intensity sound waves When such a cavity collapses, extremely high pressures are produced The process, called cavitation, is the origin of a number of mechanical, chemical and biological effects (see CAVITATION) The cavity shown in Fig. 3 began as a small hole about 10^{-5} cm in diameter and grew in size under the sound pressure. It was created by a focused barium titanate projector.

The cavitation effect can be used to disperse metals and sulfur in solutions, to produce extra-fine grain photographic emulsions, and to achieve a finer texture (smaller grain size) and more uniform alloying of a molten metal In chemistry, cavitation can be used to break long-chain polymers into shorter chains, affording a polymer of more uniform chain length than is possible with other depolymerizing methods Cavitation forces also can be used to sterilize milk.

Other high-amplitude effects. Ultrasound is used widely in the cleaning of metal parts, such as in watches. The large acoustic forces actually break off particles and contaminants from metal surfaces. Ultrasound has been investigated for the washing of textiles.

One of the principal applications of ultrasonics to gases is particle agglomeration. This depends upon the fact that light particles can follow the rapid motion of the sound waves, whereas heavy ones cannot Hence, light particles will strike and stick to heavy ones, reducing the number of small particles in the gas The heavy particles eventually will fall to a collecting plate or can be drawn there by means of an electric field This technique has been used in industry to collect fumes, dust, sulfuric acid mist, carbon black, and other substances.

Another industrial use of ultrasonics has been to produce alloys, such as lead-aluminum and lead-tin zinc, that could not be produced by conventional metallurgical techniques. Shaking by ultrasonic means causes lead, tin, and zinc to mix

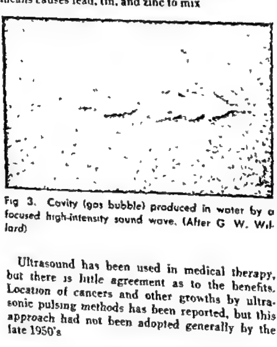


Fig 3. Cavity (gas bubble) produced in water by a focused high-intensity sound wave. (After G W. Willard)

Ultrasound has been used in medical therapy, but there is little agreement as to the benefits. Location of cancers and other growths by ultrasonic pulsing methods has been reported, but this approach had not been adopted generally by the late 1950's

ANALYTICAL USES

In addition to their engineering applications, high-frequency sound waves have been used to determine the specific types of motions that can

Ultrasonics

The science of sound waves having frequencies above the audible range, that is, above about 20,000 cycles per second (cps). Original workers in this field adopted the term supersonics. However, this name was also used in the study of air flow for velocities faster than the speed of sound. Present convention is to use the term ultrasonics as defined above. (The term silent sound also has been used to denote ultrasonic waves.) Since there is no marked distinction between the propagation and the uses of sound waves above and below 20,000 cps, the division is rather artificial. In this article, the emphasis is on instrumentation, engineering applications, and analytical uses.

ULTRASONIC PROJECTORS AND DETECTORS

The earliest instruments for producing ultrasonic waves in air were the Galton whistle and the Hartmann generator. These devices produce sound waves by blowing a jet of high pressure air from a narrow slit against a sharp metal edge The Hartmann generator raises the velocity of the jet above that of the sound waves and in effect generates standing shock waves.

Piezoelectric and magnetostriction generators. The usual types of generators for air, liquids, and solids are the piezoelectric and magnetostrictive generators. X-cut quartz crystals are used to produce longitudinal waves in gases, liquids, and solids Y-cut quartz crystals are used to produce shear or transverse waves in solids Crystals of these types are utilized in such instruments as the acoustic interferometer, which operates by sending an ultrasonic beam through a gas or liquid to be measured, and obtaining a standing wave system between the driving crystal and a reflector whose distance from the crystal can be varied by a screw system (*see* INTERFEROMETER, ACOUSTIC) Such systems provide accurate velocity and attenuation data for gases and liquids

Pulse systems. These have been used to measure properties of liquids and solids. A short burst of ultrasonic waves is sent into the medium and is reflected back. By timing the received pulse with respect to the transmitted pulse, or by a phasing technique, accurate velocity measurements can be made. Such techniques have been used widely in measuring the elastic constants of small specimens The attenuation also can be measured by the rate at which pulses decrease with distance transmitted, but careful consideration must be given to spreading loss and to the losses in the seals connecting the transducers to the specimens.

High-power devices. For higher powers, ferroelectric ceramics, such as barium titanate (Fig 1), or magnetostrictive materials, such as nickel or ferrites, commonly are employed. They generally are used for such things as ultrasonic cutting and wear or fatigue testing.

Shear waves in liquids. A number of shear wave transducers, most of them employing torsional or shear wave generators of quartz, have been used to measure the shear viscosity and shear stiffness of liquids. By this means, it has been proved that moderately viscous liquids have elastic as well as viscous properties. The viscoelastic properties of lubricating oils have been shown to contribute to the load-carrying capacity of spur gears operating at high speeds

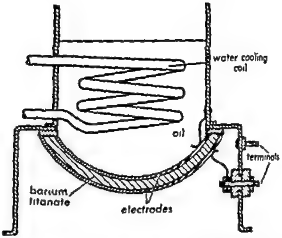


Fig. 1. Cross-sectional view of a barium titanate ultrasonic generator. The barium titanate is made to generate sound waves. The material to be treated is placed in a glass container and lowered into the oil. (From H F Olson, Acoustical Engineering, 3d ed, Van Nostrand, 1957)

ENGINEERING APPLICATIONS

The engineering applications of ultrasonics can be divided into those dealing with low-amplitude sound waves and those dealing with high-amplitude waves.

Low-amplitude applications. Low-amplitude applications are in sonar (an underwater-detection apparatus), in the measurement of the elastic constants of gases, liquids, and solids by a determination of the velocity of propagation of sound waves, and in a number of ultrasonic devices such as delay lines, mechanical filters, inspectoscopes, and thickness gages For a detailed discussion of sonar, see SONAR.

All of these applications depend on the modifications that boundaries and imperfections in the materials cause in wave-propagation properties The attenuation and scattering of the sound in the media are important factors in determining the frequencies used and the sizes of the pieces that can be utilized or investigated.

Delay lines These are useful for storing information for a certain period of time Such lines are used in moving-target indicator radar systems, in pulse-decoding systems, and in computers. *See* DELAY LINE.

Mechanical filters These are used for separating telephone communications sent simultaneously over one transmission line Figure 2 shows the construction of a torsional wave filter and the resulting attenuation characteristic. The filter consists of a number of half-wave torsic [illegible] ission lines

the grains in a polycrystalline material measured in centimeters The factor R multiplying the frequency term may vary from 0 to 10^{-3}. This source produces an appreciable internal friction for materials having a large anisotropy and a uniform grain size

Other relaxations. A number of relaxation-phenomena are associated with the motion of impurity atoms, grain boundaries, domain boundaries, and other motions occurring in a solid Interstitial atoms, such as nitrogen and carbon in iron, can cause an appreciable acoustic loss. These impurity atoms have preferred positions between the iron atoms in the crystal lattice. When a sound wave stretches the lattice in one direction and compresses it in a direction perpendicular to the first, the interstitial atoms, actuated by thermal energy, tend to go to the most open regions When a compression due to the sound wave occurs, the reverse motion takes

energy

quency

$$f = f_0 e^{-H/RT} \tag{7}$$

where f_0 is frequency of vibration of nitrogen atom due to thermal motion ($\cong 10^{12}$ cps), R, energy necessary to increase temperature of 1 mole of atoms (6.025×10^{23} atoms) by 1°C; and T, temperature in degrees Kelvin (°K) Since H is about 16,400 cal/mole for nitrogen and R is 2 cal, the relaxation frequency for this process is about 1 cps at room temperature. Other relaxations involving substitutional atoms have been observed at higher temperatures since the substitutional atoms in this case have higher activation energies Relaxations involving the rotation of grains in polycrystalline samples have been observed at high temperatures and low frequencies

Much faster relaxations occur in magnetic processes involving the motion of domain walls in magnetic materials. A demagnetized specimen is made up of a number of domains within which the direction of magnetism is the same. Domains with directions of magnetism at right angles to or at 180° from the original direction are separated by regions called Bloch walls in which the direction of magnetism changes from one domain to the other by small steps in the orientation of magnetism (For a discussion of Bloch walls, *see* FERROMAGNETISM.) A compressive stress in the same direction as the magnetic flux—for a positive magnetostrictive material—causes the domain to shrink, while it causes domains directed at 90° to expand Hence, the domain wall moves as the stress changes from compressive to extensional As the direction of magnetism changes, eddy currents are generated These limit the velocity with which a domain wall can move For a given size domain, there is some frequency for which the velocity is only half as large as that for low frequencies for the same applied magnetic field The loss at this frequency will be a maximum, and hence, this frequency is a relaxation frequency. It can be shown that this frequency is determined by the equation

$$f_0 = \frac{R}{96\chi_0 l^2} \tag{8}$$

where R is the electrical resistivity of the material, χ_0 the initial magnetic susceptibility for a demagnetized material, and l the thickness of a domain For nickel, for example, this frequency is in the order of 10^5 cps.

Many other relaxations occur, depending on the nature of the solid state motion that can occur in the material Ultrasonic measurements carried over wide frequency and temperature ranges are powerful tools for investigating such motions.

Low temperature data. Ultrasonic waves have provided significant information on processes occurring at temperatures near absolute zero In liquids, the most important results have been obtained for liquid helium, while for solids, results have been obtained with metals at low temperatures which reveal a considerable amount of information about the mechanism of superconductivity.

Liquid helium When helium is liquefied at its

The first leads to a small acoustic attenuation for normal sound, while the second leads to the capability of transmitting thermal waves, the so-called second sound. Second sound can be initiated and detected by thermal means, and it has been found that the velocity is zero at 2.2°K, rises to a maximum of 20 meters per second (m/sec) at 1.7°K, and decreases thereafter at lower temperatures The velocity of normal sound varies from 230 m/sec near absolute zero to 180 m/sec near 2.2°K. For a detailed discussion of liquid helium, *see* HELIUM, LIQUID

Attenuation at low temperatures. At very low temperatures, the ultrasonic attenuation of pure, normally conducting metals becomes high Figure 5

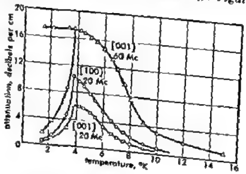


Fig. 5 Longitudinal sound-wave attenuation measurements for a single crystal of tin along the (001) axis and along the (100) axis

occur in gaseous, liquid, and solid media. Both velocity and attenuation of the sound wave are functions of the sound frequency. By studying the changes in these properties with changes of frequency, temperature, and pressure, one can obtain indications of the motions taking place.

Sound attenuation in fluids. In monatomic gases, liquefied gases, and monatomic liquids such as mercury, the sound attenuation can be explained as absorption due to viscosity and heat conduction. For such fluids, the attenuation A satisfies an equation of the form

$$A = \frac{2\pi^2 f^2}{\rho v^3}\left[\frac{4}{3}\eta + \frac{(\nu - 1)K}{C_p}\right] \qquad (1)$$

where f is the frequency, ρ the density, v the sound velocity, η the coefficient of viscosity, ν the ratio of specific heats, K the thermal conductivity, and C_p the specific heat at constant pressure.

Polyatomic liquids show additional attenuation due to relaxations of two types.

Thermal relaxations, which have been demonstrated for gases and nonassociated liquids, that is, liquids which contain nonpolar molecules, occur by an interchange of energy between the longitudinal sound wave and the rotational and internal modes of motion of the gas or liquid molecules.

Structural relaxations occur for associated liquids, for polymer liquids, and also for solids. These relaxations take place when one part of the molecule moves from one position to another under the combined effect of the thermal- and sound-wave energy. A definite structure, such as that which occurs in associated liquids and polymer liquids, is required.

Effects in solids. For solids, a variety of effects cause attenuation and velocity dispersion. Probably the simplest of these are thermal effects.

Thermal effects. When a solid body is compressed by an acoustic wave, the compressed part becomes hotter and the expanded part cooler. Thermal energy is transmitted from the hot part to the cool part. Since this energy comes from the acoustic wave, there results a loss or attenuation of the wave proportional to the square of its frequency. For bars in flexural vibration, the thermal path is quite short, and the effect produced is large. Below a frequency f_0 determined by

$$f_0 = \frac{\pi K}{2C_p\rho W^2} \qquad (2a)$$

such a source produces an internal friction $1/Q$ given by

$$\frac{1}{Q} = \frac{Y_0^\sigma - Y_0^\theta}{Y_0^\theta}\left(\frac{ff_0}{f^2 + f_0^2}\right) \qquad (2b)$$

In these equations, K is the thermal conductivity, C_p the specific heat per gram at constant pressure, f the frequency of the sound wave, ρ the density, W the width of the bar in centimeters, and Y_0^σ and Y_0^θ the adiabatic and isothermal values of Young's

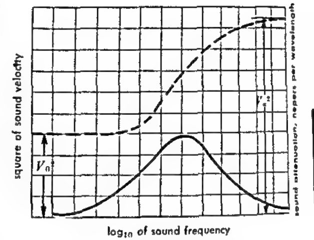


Fig. 4. Velocity dispersion (dashed curve) and corresponding attenuation per wavelength peak (solid curve) for a medium having a single relaxation

modulus, respectively. The velocity increases as a function of frequency, as shown by Fig. 4, while a corresponding internal friction occurs, as shown by the dashed line.

The thermal path l for a longitudinal wave becomes smaller as the frequency of vibration increases. It is given by the equation

$$l = v/2f \qquad (3)$$

where v is the velocity of propagation. Inserting this expression in Eq. (2a), with $l = W$, the relaxation frequency for a longitudinal wave becomes

$$f_0 = \frac{C_p\rho v^2}{2\pi K} \qquad (4)$$

Above this frequency, the material is isothermal, whereas below f_0, it is adiabatic. Inserting this value of f_0 in Eq. (2b), the internal friction due to thermal loss becomes

$$\frac{1}{Q} = \frac{Y_0^\sigma - Y_0^\theta}{Y_0^\theta}\left(\frac{f}{f_0}\right) \qquad (5)$$

for frequencies below f_0, which, from Eq. (3), is found to be usually above 10^{10} cycles and hence beyond present-day ultrasonic techniques. For other vibrational modes (longitudinal thickness), the elastic constant Y_0 is replaced by the appropriate elastic constant $\lambda + 2\mu$, where λ and μ are known as the Lamé constants. Since no compression occurs for a shear wave, it suffers no thermoelastic attenuation.

In polycrystalline materials, there is another source of thermal loss. It results from the fact that different grains are not oriented in the same way and develop different temperatures on compression. C. Zener has shown that the thermal flow produces an internal friction (relaxation loss) given by

$$\frac{1}{Q} = \left(\frac{C_p - C_v}{C_v}\right) R\,\frac{ff_0}{f^2 + f_0^2} \qquad f_0 = \frac{K}{\rho C_p L_c^2} \qquad (6)$$

where C_v is the specific heat per gram at constant volume, R is a dimensionless factor that depends on the anisotropy of the material, and L_c is the size of

prevention and cure of rickets, and the disinfection of air, water, and other substances. See ULTRAVIOLET RADIATION (BIOLOGY).

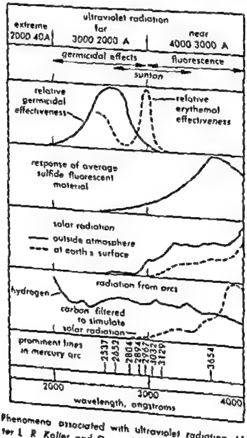
Fluorescence and phosphorescence are phenomena often generated as a result of the absorption of ultraviolet radiation. These phenomena are utilized in fluorescent lamps, in fluorescent dyes and pigments, in ultraviolet photography, and in phosphors The effectiveness of ultraviolet radiation in generating fluorescence is shown in the figure See FLUORESCENCE; FLUORESCENT LAMP; PHOSPHORESCENCE, PHOTOGRAPHY

Chemical analysis may be based on characteristic absorption of ultraviolet radiation Alternatively, the fluorescence arising from absorption in the ultraviolet region may itself be analyzed or ob- [illegible]

[illegible]

(although much solar ultraviolet radiation is absorbed in the atmosphere), arcs of elements such as carbon, hydrogen, and mercury, and incandescent bodies. The wavelengths produced by some sources of ultraviolet radiation are indicated in the figure. See ULTRAVIOLET LAMP

Artificial sources of ultraviolet light are often used to simulate the effects of solar ultraviolet radiation in the study of the deterioration of materials on exposure to sunlight Trace amounts of



Phenomena associated with ultraviolet radiation (After L R Koller and General Electric)

chemicals which strongly absorb ultraviolet radiation may effectively stabilize materials against such degradation. See INHIBITOR (CHEMICAL).

Detectors of ultraviolet radiation include biological and chemical systems (the skin, the eye of an infant, or eye without a lens, and photographic materials are sensitive to this radiation), but more useful are physical detectors such as phototubes, photovoltaic or photoconductive cells, or radiometric devices [F.W.B.]

Bibliography: L. R. Koller, *Ultraviolet Radiation*, 1952.

Ultraviolet radiation (biology)

The ultraviolet portion of the spectrum includes all radiations from 150 to 3900 A (others extend this range to 40–4000 A). Radiations shorter than 2000 A are absorbed by most substances, even by air; therefore, they are technically difficult to use in biological experimentation Radiations between 2000 and 3000 A are selectively absorbed by organic matter, and produce the best known effects of ultraviolet radiations on organisms. Radiations between 3000 and 3900 A are relatively little absorbed and are less active on organisms. Ultraviolet radiations, in contrast to x rays, do not penetrate far into larger organisms; therefore, the effects they produce are surface effects, such as sunburn and development of D vitamins from precursors present in skin or fur. The effects of ultraviolet radiations on life have, therefore, been assayed chiefly with unicell[illegible] proto[illegible] organ[illegible] have been useful as well

Ultraviolet radiation in sunlight at the surface of the earth is restricted to the span from about 2870 to 3900 A, although shorter wavelengths are present beyond our atmosphere, as shown by measurements with rockets Consequently, artificial sources of the radiations are generally used in experimentation.

Photobiological effects. Only the ultraviolet radiations which are absorbed can produce photobiological action. All life activities are shown to be affected by ultraviolet radiations, the effect depending upon the dosage. Small dosages activate unfertilized eggs of marine animals, reduce the rate of cell division, decrease the synthesis of nucleic acid, especially in the nucleus, reduce the motility of cilia and of contractile vacuoles, and sensitize cells to heat Large dosages increase the permeability of cells to various substances, inhibit most synthetic processes, produce mutations, stop division of cells, decrease the rate of respiration, and may even disrupt cells The [illegible] diations up[illegible]

Despite t[illegible] tions are u[illegible] in biological research because they stop certain cell activities [illegible] without [illegible] have [illegible] of n[illegible]

[illegible] MUTATION

shows measurements of pure tin for two directions in the crystal and for two frequencies Above 10°K, the ultrasonic attenuation is relatively small and increases as the square of the frequency. At 4°K, at which tin is still in the normal state, the attenuation is high and increases in proportion to the frequency. It has been shown that the added attenuation in the normal state is due to the transfer of momentum and energy from the acoustic wave to the free electrons in the metal. If the acoustic wavelength is greater than the electronic mean free path, this transfer determines an effective viscosity, and the attenuation increases in proportion to the square of the frequency. When the mean free path becomes longer than the acoustic wavelength, as it does at low temperatures, the energy communicated to the electrons is not returned to the acoustic wave and a high attenuation results. The attenuation is proportional to the number of times the crystal lattice vibrates and hence to the frequency.

As the temperature drops below the temperature at which tin becomes superconductive (3.71°K), this source of attenuation drops rapidly to zero. The form of the curve has been used to confirm the J. Bardeen, L. N. Cooper, and J R Schrieffer energy-gap theory of superconductivity. *See* SUPERCONDUCTIVITY; *see also* SOUND. [W.P M.]

Bibliography: L. Bergmann, *Der Ultraschall und seine Anwendung in Wissenschaft und Technik*, 6th ed., 1954; W. P Mason, *Physical Acoustics and the Properties of Solids*, 1958; E. G Richardson, *Ultrasonic Physics*, 1952; P. Vigoureux, *Ultrasonics*, 1950.

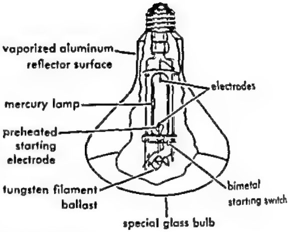
Ultraviolet lamp

A mercury-vapor lamp designed to produce ultraviolet radiation, widely used in commercial and industrial environments. Also, some fluorescent lamps and mercury-vapor lamps that produce light are used for ultraviolet effects. *See* FLUORESCENT LAMP, MERCURY-VAPOR LAMP.

Near-ultraviolet lamps. Ultraviolet energy in the wavelength region from 3200 to 4000 angstrom units (A) is known as near ultraviolet, or black light. Fluorescent and mercury lamps can be filtered so that visible energy is absorbed and emission is primarily in the black-light spectrum. The ultraviolet energy emitted is used to excite fluorescent pigments in paints, dyes, or natural materials to produce dramatic effects in advertising, decoration, and the theater, in industrial inspection, fluorescent effects are often used to detect flaws in machined parts and other products, as well as invisible laundry marks.

Middle-ultraviolet lamps. Middle ultraviolet spans the wavelength band from 2800 to 3200 A. Mercury-vapor lamps are sometimes designed with pressures that produce maximum radiation in this region, using special glass bulbs that freely transmit this energy.

One such lamp type is the sunlamp. Illustrated here is the reflector sunlamp, with a self-contained filament ballast and starting mechanism. The re-



Ultraviolet lamp.

flector sunlamp combines the middle ultraviolet, which reddens the skin, with infrared energy and light from the filament to produce a sun tanning effect with the sensations of warmth and brightness normally associated with sunshine

Other lamps designed for middle-ultraviolet radiation are known as photochemical lamps. They are used for such varied tasks as mold destruction, inspection of sheet metal for pinholes, and black and-white printing of engineering drawings.

Far-ultraviolet lamps. Some radiation in the 2200- to 2800-A wavelength band has the capacity to destroy certain kinds of bacteria. Mercury lamps designed to produce energy in this region (the 2537-A mercury line) are electrically identical with fluorescent lamps; they differ from fluorescent lamps in the absence of a phosphor coating and in the use of glass tubes that transmit far ultraviolet Germicidal lamps are sometimes used to reduce the airborne bacteria count, and to kill certain organisms on or near the surface of perishable products in storage, or certain products in the pharmaceutical industry. *See* ULTRAVIOLET RADIATION (BIOLOGY) [A MA]

Ultraviolet radiation

Electromagnetic radiation in the wavelength range 40–4000 A The ultraviolet region begins at the short wavelength (violet) limit of visibility and extends to the wavelength of long x-rays. It is loosely divided into the near (4000–3000 A), far (3000–2000 A), and extreme (below 2000 A) ultraviolet regions (see figure). In the extreme ultraviolet, strong absorption of the radiation by air requires the use of evacuated apparatus, hence, this region is called the vacuum ultraviolet. Important phenomena associated with ultraviolet radiation include biological effects and applications, the generation of fluorescence, and chemical analysis through characteristic absorption or fluorescence.

Biological effects of ultraviolet radiation include erythema or sunburn, pigmentation or tanning, and germicidal action. The wavelength regions responsible for these effects are indicated in the figure Important biological uses of ultraviolet radiation include therapy, the production of vitamin D, the

important in research than in clinical practice at the present time. *See* RADIATION BIOLOGY; TUBERCULOSIS; VITAMIN D [A.C.GI.]

Bibliography: A. C. Giese, Ultraviolet radiations and life, *Physiol. Zool.*, 18:223–250, 1945; A. Hollaender (ed.), *Radiation Biology*, vol. 2, 1955; J. Jagger, Photoreactivation, *Bacteriol. Rev.*, 22:99–142, 1958.

Umbellales

An order of the plant subclass Dicotyledoneae, characterized by flowers borne in determinate umbels. There are 3 families: the ginseng family (Araliaceae) with 65 genera and 800 species, mainly tropical; the carrot family (Umbelliferae) with 125 genera and 2900 species, and the dogwood family (Cornaceae) with 18 genera and 125 species. Ginseng, English ivy, and the ornamental aralias belong to the ginseng family. The Cornaceae include the dogwoods, much planted as ornamentals, and the *tupelo*, a valuable timber tree. Among the Umbelliferae are the food plants celery, parsley, fennel, parsnip, and carrot. Species which yield choice flavoring materials are coriander, cumin, caraway, dill, and anise. Myrrh (*Myrrhis odorata*) is grown for its pleasing odor. Poison hemlock (*Conium maculatum*) and water hemlock (*Cicuta maculata*) are very poisonous plants. See separate articles for trees, food plants, and flavoring materials listed in this article. *See also* DICOTYLEDONEAE; EMBRYOPHYTA, PLANT KINGDOM.

[P.D.S.]

Umbra

That portion of a shadow which is screened from light rays emanating from any part of an extended, or broad, source. With a point source, the entire shadow consists of an umbra, since there can be no region in which only part of the source is eclipsed. If the source has an appreciable extent, however, there exists a *transition* region surrounding the umbra called the [illegible]
part o [illegible]
bra is [illegible]
source [illegible]
SHADO [illegible]
NOMICAL, PENUMBRA, [illegible] ASTRO-

[F.A.J.]

Umklapp process

A concept in the [illegible]
solids whic [illegible]
or more wa [illegible]
electron w [illegible] a continuum, such interactions occur only among waves described by wave vectors $\mathbf{k}_1$, $\mathbf{k}_2$, and so on such that the interference condition

$$\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0 \tag{1}$$

is satisfied. The sign of k depends on whether the wave absorbs or emits energy. Since $\hbar k$ is the momentum of a quantum (or particle) described by the wave, Eq. (1) corresponds to conservation of momentum. In a crystal lattice further interactions occur, satisfying

$$\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = \mathbf{b} \tag{2}$$

where $\mathbf{b}$ is any integral combination of the three inverse lattice vectors $\mathbf{b}_i$, defined by $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, the a's being the periodicity vectors (*see* CRYSTAL STRUCTURE). The group of processes described by Eq. (2) are the Umklapp processes or flip-over processes, so called because the total momentum of the initial particles or quanta is reversed.

Examples of Umklapp processes are (1) interactions of three lattice waves due to anharmonic lattice forces; of these, only processes described by Eq. (2) produce intrinsic thermal resistance in nonmetals; the exponential variation of the thermal resistance observed in dielectric crystals at low temperatures confirms the concept of Umklapp processes; (2) scattering of *electrons* by lattice waves, causing electrical and thermal resistance in metals; it has become clear that the observed properties cannot be accounted for in terms of processes described only by Eq. (1), but Umklapp processes must also be considered; (3) *Bragg reflection*, which can be regarded as an Umklapp process involving only two waves. *See* CONDUCTION (HEAT); ELECTRICAL CONDUCTIVITY OF METALS; X-RAY DIFFRACTION.

[P.G.K.]

Bibliography. C. Kittel, *Introduction to Solid State Physics*, 2d ed., 1956.

Uncertainty principle

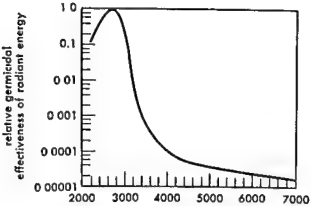
In quantum mechanics the precept (W. Heisenberg, 1927) that accurate measurement of an observable quantity necessarily produces uncertainties in one's knowledge of the values of other observables. It is also called the indeterminacy principle.

In particular, for a single particle

$$\Delta x\, \Delta p_x \gtrsim h/2\pi \tag{1a}$$
$$\Delta t\, \Delta E \gtrsim h/2\pi \tag{1b}$$

In Eq. (1a), Δx represents the uncertainty (error) in the location of the x coordinate of the particle at any instant, and Δp_x is the simultaneous uncertainty in its x component of momentum; h = Planck's constant = 6.61×10^{-27} erg-sec. Equation (1b) is more subtle than Eq. (1a), and its interpretation depends somewhat on the circumstances. For instance, Eq. (1b) relates the uncertainty ΔE in an energy measurement to the time interval Δt during which the measurement was performed; however, Eq. (1b) also relates the uncertainty ΔE in the energy radiated by a system to the uncertainty Δt in the time at which it radiates, that is, the uncertainty in its lifetime. For derivation of Eqs. (1) from the formal postulates of quantum theory, see QUANTUM THEORY, NONRELATIVISTIC. For a discussion of the physical significance of Eqs. (1), especially their relation to the dual wave-particle properties of matter and radiation (summarized in the *de Broglie* relations $E = hf$, $p = h/\lambda$, where f is the frequency and λ the wave-

Action spectra. Some wavelengths of ultraviolet radiations are more effective than others. More bacteria are killed by a given dosage of ultraviolet radiation at 2600 A than by the same dosage of radiation at 3000 A. When the bactericidal effectiveness of each of a series of wavelengths is plotted against the wavelength, the resulting curve is an action spectrum for the bactericidal effect. Action spectra have been determined for many other effects of ultraviolet radiations. Each action spectrum is postulated to represent the absorption spectrum of the substance in the cell responsible for the particular effect For the bactericidal effect, production of mutations, and retardation of cell division, the action spectrum suggests absorption by nucleoproteins or nucleic acid. For sensitization to heat and inhibition of ciliary movement, the action spectrum suggests absorption by ordinary proteins For permeability, still another action spectrum exists.



Relative bactericidal action of the near-ultraviolet and visible regions for *Escherichia coli* on agar (*From A. Hollaender, ed., Radiation Biology, vol. 2, McGraw-Hill, 1955*)

Mechanism. The effect of a given dosage of ultraviolet radiation upon cells is greater when the radiation is flashed than when continuous, that is, if a period of radiation is followed by a period of darkness. This seems to indicate that a thermal reaction follows the primary photochemical reaction, a suggestion which is substantiated by the fact that increasing the temperature within the viable range accentuates the effect of flashing

Photoreversal. The action of ultraviolet radiation on cells can be reversed to a considerable degree by simultaneous or subsequent exposure of the irradiated cells to short wavelength visible, violet and blue, or long wavelength ultraviolet light. This process has been called photoreversal or photoreactivation. Thus, nucleic acid synthesis, inhibited by ultraviolet radiations, is resumed after exposure to visible light At the same time, cell division, previously inhibited or retarded, is resumed. It appears that those effects of ultraviolet radiation having a nuclear site are most readily photoreversed

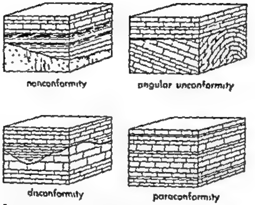
Photoreversal is never complete; therefore, photoreactivated cells act as if they had been given a smaller dosage of ultraviolet radiations It is evident that to make most effective use of ultraviolet radiations as a tool in experimental work, cells must be protected from visible light.

Effects of ultraviolet on the skin. Erythema is the reddening of the skin following exposure to ultraviolet radiation of wavelength shorter than 3200 A, wavelength 2967 A being most effective These radiations injure cells in the outer layer of the skin, or epidermis, liberating substances which diffuse to the inner layer of the skin, or dermis, causing enlargement of the small blood vessels A minimal erythemal dose just induces reddening of the skin observed 10 hours after exposure. A dose several times the minimal gives a sunburn, killing some cells in the epidermis after which serum and white blood cells accumulate, causing a blister. After the dried blister peels, the epidermis is temporarily thickened and pigment develops in the lower layers of the epidermis, both of these factors serving to protect against subsequent exposure to ultraviolet. Both thickening of the epidermis and tanning may occur without blistering. Since the pigment in light-skinned races develops chiefly below the sensitive cells in the epidermis, it is not as effective as in dark-skinned races where the pigment is scattered throughout the epidermis Consequently, the minimal erythermal dose is much higher for the dark- than for the light-skinned races.

Pigmentation or tanning also appears when the skin of young individuals is subjected to massive doses of ultraviolet radiations longer than 3200 A Presumably this occurs by oxidation of precursors of the pigment, melanin, already present in the epidermis. Since such radiation is strong in sunlight, a skin may tan even if the short radiations are excluded

Excessive exposure to ultraviolet radiation has been found to lead to cancer in mice, and it is claimed by some to cause cutaneous cancer in man

Clinical use. Ultraviolet radiations were once used extensively in the treatment of rickets, many skin diseases, tuberculosis other than pulmonary, especially skin tuberculosis (lupus vulgaris), and of many other diseases The enthusiasm for skin bathing is, in part, a relic of the former importance of ultraviolet radiation as a clinical tool. Vitamin preparations, synthetic drugs, and antibiotics have either displaced ultraviolet radiations in such therapy or are used in conjunction with the radiations Ultraviolet radiations alone are still employed to treat rickets in individuals sensitive to vitamin D preparations. In conjunction with chemicals, they are used in treating skin diseases, for example, psoriasis, pityriasis rosea, and sometimes acne, as well as for the rare cases of sensitivity to visible light. They are also often used to sterilize air in hospitals. In some European laboratories, they are still used as adjuncts to drugs for treating lupus vulgaris and some other forms of tuberculosis Ultraviolet radiations, however, are probably more



Four types of unconformity (From C. O. Dunbar and J. Rodgers, Principles of Stratigraphy, Wiley, 1957)

although the names applied to each are not universally agreed upon:

1. Nonconformity—rocks below the break are not stratified, such as massive crystalline rocks.
2. *Angular unconformity*—rocks below the break are stratified but lie at an angle to those above.
3. Disconformity—strata below are parallel to those above but are separated by an evident surface of erosion.
4. *Pseudoconformity* or paraconformity—strata are parallel but surface is hardly distinguishable from a simple bedding plane.

Importance in geologic record. Unconformities are commonly the chief or only record in a given stratigraphic sequence of the time elapsed between the deposition of the rocks above and below. This time is called the hiatus. Unconformities are also important in deciphering the events that took place during that time, such as erosion, deformation, metamorphism or igneous intrusion. Nonconformable strata often point to major crustal disturbances preceding the erosion represented by the unconformity. The geologic date of the erosion interval and of preceding deformation may be somewhat narrowly defined if there is little difference in the age of the oldest rocks above and youngest rocks below the unconformable surface. *See* STRATIGRAPHY. [J.R.]

Underwater demolition

The destruction or fragmentation of underwater obstacles by the use of explosive charges hand-placed by diver personnel. The techniques of underwater demolition were thoroughly developed during World War II. Underwater demolition is primarily employed as a short range, emergency measure to accomplish a military object ... of its ... is the ... range ...

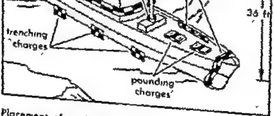
Equipment used. Diving equipments used by naval personnel in underwater demolition include deep-sea equipment, light-weight equipment, and SCUBA (self contained underwater breathing apparatus) gear. Standard demolition equipment and techniques are employed. The explosive requirement for a specific task would be judged by the properties of the individual high explosive (rapidity of detonation, sensitivity, brisance, and ...

various diving programs, that is, salvage, explosive ordnance disposal, and underwater demolition teams. Each program, operating within the general framework of diving and demolition procedure, is responsible for specialized functions and emphasizes those functions in its individual training requirements. For example, underwater demolition teams (frogmen), primarily used offensively in amphibious landings, have developed distinct procedures for demolishing natural and man-made obstructions which block the water approaches to a proposed assault beach. Likewise, the explosive ordnance disposal diver must be capable of performing specific underwater demolition procedures in a mine clearance operation.

The application of underwater demolition procedures is probably most varied in the salvage forces of the U.S. Navy. The following discussion, involving a typical salvage operation, describes many of these procedures.

Consider a merchant ship that is sunk, blocking a harbor entrance to the extent that waterborne movement into or out of the harbor is denied. The required depth to allow harbor entrance over the sunken vessel is 36 ft. Main deck level of the vessel is at 26 ft, requiring either refloating or ship flattening to gain the 36-ft depth. Diver reconnaissance establishes that hull damage to the vessel would require extensive patching. In order to allow immediate harbor access, the decision is made to flatten the vessel by employing demolition procedures.

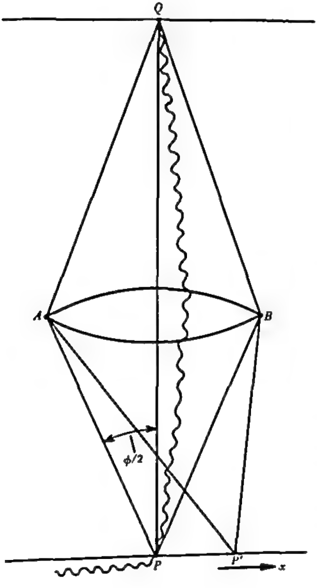
In the case of the diagrammed vessel, the follow-



Placement of explosive charges on a sunken merchant ship blocking a harbor

length), *see* QUANTUM MECHANICS. The present article deduces Eqs. (1) from analyses of specific experiments. Only two simple experiments are examined here. Although the sources of error are sometimes arcane, no actual or imagined (*Gedanken*) experiment has failed to agree with the restrictions of Eqs. (1).

Position measurement. In the figure, an electron is viewed with light of wavelength λ A photon scatters from the electron, through the lens, and is observed at Q. If the microscope had infinite resolution, observing a photon at Q would precisely locate the electron at P; with the actual lens of finite aperture AB, a photon has an appreciable probability of reaching Q if the electron is as far from P as P', where $P'A - P'B \cong \lambda/2$. Thus, the uncertainty in position of the electron is $\Delta x \cong 2PP' \cong \lambda/\sin\phi$, where ϕ is the aperture angle of the lens (at P) The photon, which originally had momentum h/λ in the x direction, may have x momentum anywhere between $\pm(h/\lambda)\ \sin(\phi/2)$ after scattering, depending on the path the photon



Microscope for locating electron

takes through the lens. Thus, the uncertainty in x momentum transferred to the electron is

$$\Delta p_x \cong 2(h/\lambda)\ \sin(\phi/2) \cong (h/\lambda)\sin\phi$$

Therefore, $\Delta x\ \Delta p_x \cong h$, which is consistent with Eq. (1*a*). *See* RESOLVING POWER (OPTICS).

Energy measurement. Suppose the kinetic energy $E = \frac{1}{2}mv^2$ of a particle moving along x is found by measuring its speed v. Then $\Delta E = mv\ \Delta v = v\ \Delta p_x$. If the measurement takes a time Δt, the position of the particle is uncertain by an amount $\Delta x = v\ \Delta t$. Then $\Delta E\ \Delta t = \Delta p_x\ \Delta x$ which, accepting Eq. (1*a*), demonstrates Eq (1*b*).

The uncertainty principle refers only to uncertainties at the instant of measurement, which uncertainties prevent complete certainty about the future course of the system under observation. Measurements made at time t_0 may reveal the precise position and momentum a particle had at a time $t < t_0$; this information is not usable for predictions at times $t > t_0$ however, once the past revealing but future-obscuring measurement at t_0 has been performed. Eqs. (1), which are deduced nonrelativistically, are modified when the special theory of relativity is taken into account (*see* RELATIVISTIC MECHANICS; RELATIVITY). For example, relativity requires the recoil photon in the figure to have a longer wavelength than the incident photon, with the result that Δx cannot be less than $\sim h/mc$, Eq (1*a*) permits precise position measurements, $\Delta x \rightarrow 0$, provided $\Delta p_x \rightarrow \infty$ is acceptable. *See* COMPTON EFFECT; QUANTUM ELECTRODYNAMICS. [E.G]

Bibliography: W. Heisenberg, *The Physical Principles of Quantum Theory*, 1949; *see also* QUANTUM MECHANICS; QUANTUM THEORY, NONRELATIVISTIC

Unconformity

The relation between adjacent rock strata whose time of deposition was separated by a period of nondeposition or of erosion. The break in the stratigraphic sequence represents a significant change in regimen (in contrast to diastem). Thus marine strata separated by a surface of subaerial erosion are unconformable, but those separated by a surface of mere nondeposition or of erosion by shifting currents which do not represent an over all change of conditions are not. Similarly strata deposited by a river but separated by an erosion surface representing scour by that river during its normal shifting on its flood plain are not in general unconformable The word unconformity is sometimes also used to denote the surface between the unconformable strata *See* DIASTEM.

Types of unconformities. The term unconformity was originally used only where there was an obvious structural discordance between the rocks above and below the break, as where strata below had been tilted and then beveled by erosion before those above were deposited. This restricted usage is still advocated by some. In general four main kinds of unconformable rel[illegible] can be recogni[illegible]

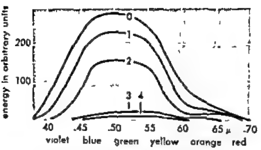


Fig 3 Energy spectra at a depth of 10 m in different types of water Curves marked 0, 1, 2, 3, and 4 represent energy spectra in pure water, clear oceanic, average oceanic, average coastal, and turbid coastal sea water, respectively. (*From H U. Sverdrup et al, The Oceans, Prentice-Hall, 1942*)

The short distances at which underwater photographs must ordinarily be taken necessitates the use of an extremely wide lens Wide angle lenses are further required because of the magnification (approximately 25%) of objects underwater, owing to the difference between the index of refraction of the water and air in the underwater camera housing Special lenses have been developed to correct for this magnification

Underwater photography equipment is available for work at almost any depth The photographer, not the equipment, is usually the limiting factor. He can remain underwater only as long as it is safe to do so and must continually be aware of his physiological limitations. In general, if he is experienced and properly trained, a diver can safely descend to 50 m, but even so, he can remain there only a relatively short time to find a subject, set up lights if needed, and take pictures. Deeper pho-

... about the maximum depth limit for handheld underwater photography. [R F DL]

Deep-sea underwater photography. Photography in the deep sea involves the design and use of camera and lighting equipment that will operate deeper than 40 fathoms (75 m) and meet the following requirements: (1) it must withstand the pressure of the water (1 atmosphere for each 32 ft), and (2) it must operate to give the desired result, such as photographs of objects in midwater or at the bottom For example, bottom switch-, sound-, and light actuated cameras have been made for special purposes.

Camera housing Cases for cameras for the deepest ocean must resist a pressure of 1100 atmospheres (17,000 psi). The optical windows and the seals for electrical connections must be carefully designed and adequately tested, or failure at sea will result. Cases of cylindrical shape (Fig 4), especially of a 4-in. inside diameter casing, have been designed to house cameras used at depths of 25 000 ft (Fig. 5) and those used on bathyscaphs. The 4-in diameter case size is just large enough to accept ... 100-ft ... of 35-mm film which holds 800 exposures. Thin Cronar film base increases the number of exposures to 1000. Although photographs can be taken through the viewing porthole of bathyscaphs by the observers, external remote-controlled cameras are preferred, since the portholes are rather small and do not permit convenient viewing when the camera is in position *See* BATHYSCAPH.

Light source. Light is required for photography in the deep oceans. The expendable flash bulb has been extensively used. The newer electronic flash light source has many advantages as an illumination system, since it is efficient, can be flashed many times, has a daylight color quality, can be controlled electrically to flash at a desired instant, and has a short flash time, thus arresting motion of the subject.

Bottom photography. Cameras used to photograph the ocean bottom are of several types: (1) the type which triggers a flash lamp when a pogo stick on the bottom of the camera touches the bottom; (2) the free-floating camera which sinks

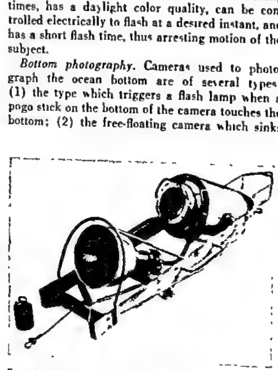


Fig 4 Deep-sea camera designed by C J Shipek. Operating depth is 20,000 ft (*U S. Navy Electronics Laboratory*)

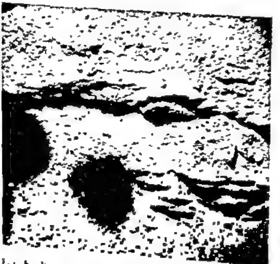


Fig 5 Photograph taken in Romanche Trench, 0°10'S, 18°21'W, at depth of 25,000 ft

ing procedure might be used. All of the explosive charges would be hand-placed by divers. First, trenching charges are laid under the hull to settle the ship. Then the masts are cleared away by cutting charges which employ a shearing action much like felling a tree. The superstructure is removed or dispersed by a combination of linear cutting charges The hull is flattened in two stages: the bow and stern are removed by linear cutting charges, then pounding charges are connected along the main deck line in combination with cutting charges at the deck edges. *See* EXPLOSION AND EXPLOSIVE; SCUBA; SHIP SALVAGE; SKIN DIVING. [H.W.P.]

Underwater photography

The technique of using photographic equipment underwater. The term also refers to the equipment used in this specialized phase of photography.

Shallow underwater photography. In shallow underwater photography, or hand-held underwater photography as it is more commonly called, the photographer, wearing some type of diving equipment, is submerged with his camera and takes photographs directly (Figs 1 and 2). This type of underwater photography did not become a technical or scientific tool until the development and widespread use of self-contained underwater breathing apparatus, or SCUBA. *See* SCUBA; SKIN DIVING.

Underwater photography is controlled by many different factors, some of which are available light, clarity of water, differential absorption of light in various wavelengths, film capability, and construction of underwater housing for cameras. The underwater photographer can never expect to take clear pictures at distances greater than 50 meters In most instances, he will have to depend on natural light for the exposure of his film. Flash bulbs and electronic flash units are often used under water as supplemental light sources. However, care must be taken when using artificially supplied light or natural light at depths greater than a few feet if true color balance is desired, because of the differential absorption of light by water. Red light energy and other warm light is absorbed at a much higher rate than the colder light waves. Light in the blue wavelengths penetrates deepest into the sea Infrared and ultraviolet wavelengths are rapidly attenuated as they travel through the water medium and are not the good sources of penetrating light energy that they are in air (Fig. 3). Filters are useful, but because of the generally low light level and actual lack of light in certain wavelengths of the light energy spectrum underwater, they have the tendency to do more harm than good

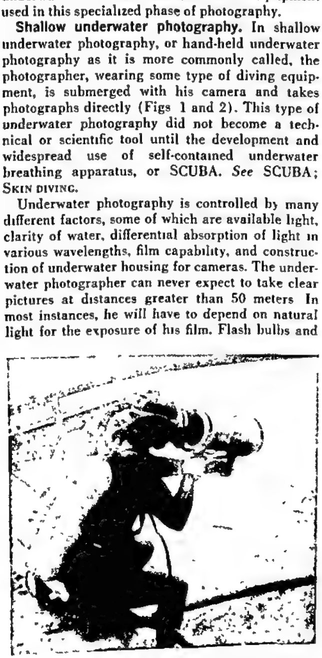


Fig 1. Photographer using underwater motion picture camera to record data on the bottom. This picture was taken at a depth of 120 ft High-speed Tri-X film was [illegible] Camera settings: $1/100$ sec and $f/4.0$.

Fig 2 Recording the action of a pressure regulator by a hand-held underwater still camera (Rolli-marine) Direct observation and photographic recording, in situ, have become extremely important tools under water

The transparency of the water is most often the limiting factor. It is dependent on the organisms and sediment content of the water and is quite variable. Contrary to popular belief, it cannot be improved by artificial light. The small particles in the water scatter the light energy, and therefore more light makes pictures worse in turbid water Water with the highest visibility is located away from the continental land masses in deep oceanic water or in areas where very little sediment is added to the sea water, such as off rocky or arid coasts. The highest visibility to be expected in the clear oceanic waters is about 100 *m* Occasionally, these clear waters lap up onto the margins of the continents and thus temporarily improve water visibility in coastal areas However, in most instances, water visibility is less than 30 m The average underwater visibility off southern California is about 5 m. For a discussion of water [illegible]r and transparency, *see* SEA WATER.

pek, *A Deep-Sea Photographic Study of Some Marine Environments in the Eastern and Southeastern Pacific*, 1959, E. M. Thorndike, Deep-sea cameras of the Lamont Observatory, *Deep-Sea Research*, 5(3):234-237, 1959

Underwater sound

The production, transmission, and reception of sounds in the ocean is an important part of modern acoustics. Since the ocean is highly absorptive to light and other electromagnetic radiations, but not to sound waves, acoustics has found many applications in probing the sea. Acoustical methods of detection, for example, are useful in locating submarines and other submerged objects, and so have many applications in naval science. Underwater sound is also important in scientific research: oceanographers and geophysicists employ sound waves to determine the physical structure of the ocean and its bottom, and sound is used by marine biologists to study organisms found in the sea.

Sound propagated between two points in the ocean is affected by many factors which vary geographically and, in a given location, may change seasonally or even diurnally. One of the most important effects is the refraction of sound by vertical variations in the velocity of sound, the latter being determined at any point by the pressure, temperature, and salinity. A sound wave is also attenuated by absorption processes in sea water, this effect being more pronounced as the frequency increases. Scattering and reflection of sound take place at the boundaries of the ocean and are also caused by suspended bubbles or marine life. Finally, the reception of man-made sounds also may be limited by the presence of interfering sounds generated by surface waves or emitted by marine life. The influence of these factors is considered in the following sections.

Velocity of sound. Sound velocity in the ocean varies roughly between 4800 feet per second (fps) and 5100 fps. At 70°F and with a normal salinity of 34 parts per thousand by weight, the sound velocity is 4935 fps at the ocean surface. With few exceptions (for example, at the edge of the Gulf Stream), the variations in the ocean's properties are almost entirely in the vertical direction.

The sound velocity increases with an increase in temperature, pressure, or salinity. Pressure increases the velocity by 1.82 fps per 100 ft in depth, and salinity increases the velocity by 4.3 fps for an increase of 1 part per thousand in salt content. The amount by which a temperature change affects the velocity depends upon the temperature: at 32°F, the increase is 8.4 fps/°F; at 70°F, it is 5.0 fps/°F.

Taking all three factors into account, the sound velocity V in feet per second can be approximated by the following equation:

$$V = 4422 + 11.25T - 0.0450T^2 + 0.0182D + 4.3S - 34$$

where T is the temperature in °F, D is the water depth in feet, and S is the salinity in parts per thousand.

Temperature is the most important factor in determining the sound velocity dependence on depth. Salinity changes are not important in the open ocean, being significant only where there is large-scale mixing of different bodies of water, for example, at a river mouth or at the Straits of Gibraltar. The increase of V with pressure is universal, depending only on the depth. The temperature-depth variation at any location can be determined by use of a bathythermograph which provides a

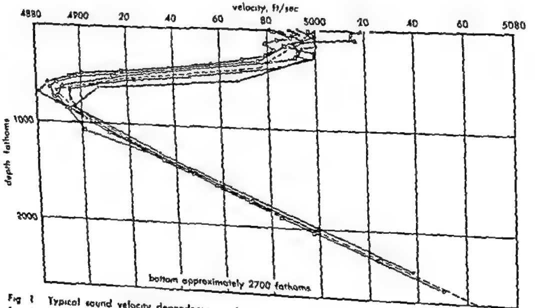


Fig 1 Typical sound velocity dependence

to the bottom, exposes a picture, and then rises to the surface when a weight is released; and (3) the type which takes many photographs at regular intervals (such as 10 seconds) and depends upon positioning with sonar or other methods of adjustment. The standard 100-ft (33-m) reel of film can take 800 double frame 35-mm photos. With expendable flash bulbs, the equipment is hauled to the surface after each photo. With electronic flash lighting, the equipment can be operated many times. Most of the deep-sea photography efforts of the future will undoubtedly use electronic flash because of the advantages previously listed.

Lowering and retrieving. The steel cable which is used to lower a camera to great depths has several disadvantages First, the weight of the cable becomes considerable; second, the winch operator cannot tell when the camera touches the bottom, since the cable is so heavy; and third, if the camera does rest on the bottom, the cable may kink, causing a break when the cable is pulled in and strikes the pulley. Cameras have been lost in this manner.

Nylon has some advantages for lowering cameras to great depths, since the material is almost weightless in the water; thus the winch operator can tell when the camera touches the bottom However, the winch problem is serious owing to the elastic properties of the rope. For example, the nylon after being pulled up should be allowed to shrink before being wound on the take-up drum.

Summary. The usefulness of underwater photography as a tool in oceanographic research (Fig. 6) has grown rapidly with advances in the design of underwater cameras and the development of photographic techniques such as stereophotography (Fig. 7) Photography is probably inferior to personal observation, although television may eventually be developed as a rival to the camera for deep-sea operation. An ideal situation is to use the television equipment as a remote view finder for the camera, thus allowing the operator on the surface to take his photograph at the opportune moment by remote observation and control. Photographic cameras obtain pictures with greater detail than is possible with present television apparatus and are not restricted to the comparatively shallow depths of such equipment *See* UNDERWATER TELEVISION.

[H.E.E.]

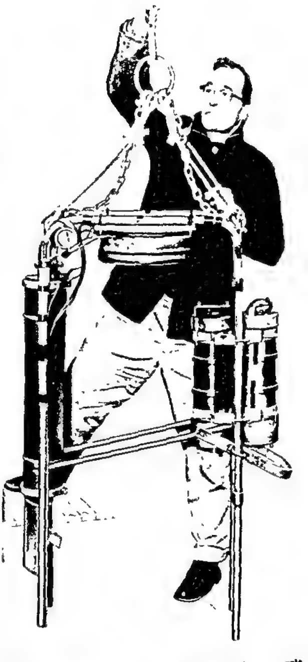


Fig 7 Deep-sea stereo camera designed to operate at depth of 36,000 ft Sonar transducer at top of camera signals to the surface both directly and by reflection from the bottom of the sea to determine the height of the camera above the bottom.

Fig 6 Photograph of Mid-Atlantic Ridge at depth of [illegible] ft (Lamont Geological Observatory)

Bibliography: H E. Edgerton and L. D Hoadley, Cameras and lights for underwater use, *J. Soc. Motion Picture Television Engrs.*, 64:345–350, 1955; A. S. Laughton, A new deep-sea underwater camera, *Deep-Sea Research*, 4(2): 120–125, 1957; D M. Owen, Deep-sea underwater photography and some recent stereoscopic applications, *Photogrammetric Engineering*, 17(1):1 ; C. J. Shi-

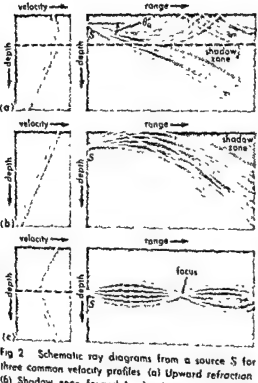


Fig 2 Schematic ray diagrams from a source S for three common velocity profiles (a) Upward *refraction* (b) Shadow zone formed by limiting ray. (c) *Sound focusing*

range can be extended only by lowering the sound source

Sound channels These are formed when there is a velocity minimum at some depth This is illustrated in Fig 2c where a source and receiver are at the depth of a velocity minimum. Rays are alternately refracted upward and downward, remaining trapped by refraction. *Propagation* is extremely favorable being better than in the surface channel, since no *reflection* is needed to confine the rays A unique property of such a channel is sound focusing A bundle of rays leaving at angles near the horizontal are brought back together periodically, the distances between foci being characteristic of the gradients At these *focal points* (in three dimensions, concentric circles), the sound intensity is very high For an example of such a sound channel see SONAR

Propagation loss. This is caused by refraction and by absorption If I_0 is the acoustic intensity 1 yard from the source of sound then the ...

... measures the effect of geometrical spreading and refraction If the source is directional, then I_0 is dependent on the direction from which the received ray was originally emitted from the source

Propagation loss measurements are usually expressed in decibel (db) units as follows. The source level S is taken as the intensity at 1 yard relative to some reference level I_r, that is, $S = 10 \log I_0/I_r$ (The reference level usually taken is that corresponding to rms pressure of 1 dyne/cm².) If $L(R)$ is the measured acoustic intensity at R, relative to the same reference level, then $L(R) = 10 \log I(R)/I_r$. The propagation loss can then be measured by $H = S - L(R)$, where H is called the transmission loss. From the preceding definitions, it is given by $H = -10 \log F(R) + \alpha R$, where α is measured in db per yard. In the ideal case where there are no reflecting boundaries and the velocity of sound is constant, $F(R) = R^{-2}$ (i.e., the inverse square law holds), and so $H = 20 \log R + \alpha R$. See INVERSE-SQUARE LAW.

Another quantitative description of the loss is the transmission anomaly A, where $A = H - 20 \log R$. Thus, A compares the propagation with that of an ideal medium where absorption, reflection, and refraction effects are absent.

Refraction in practice gives rise to widely differing propagation losses. In a sound channel, $F(R)$ is sometimes found to be much closer to R^{-1} than to R^{-2}, and in a shadow zone, $F(R)$ decreases much faster than R^{-2}.

Reflection of sound. This occurs when sound strikes the sea surface, the sea bottom, or large objects in the water. If such surfaces are flat over distances large compared to the wavelength of the sound, then a fraction of the sound is reflected at an angle with respect to the normal which is equal to the angle of incidence. A quantitative description can be approximated by considering the reflection at a boundary between two fluids; here, the fraction of reflected energy is given by

$$r = \left[\frac{\rho_1 V_1 - \rho V B}{\rho_1 V_1 + \rho V B}\right]^2$$

In this equation, the ray is incident from the medium having sound *velocity V and density ρ* upon a medium having values V_1 and ρ_1. The product ρV is called the specific acoustic impedance, and B is given by

$$B = [1 - (1 - V_1^2/V^2) \tan^2 \theta]^{1/2}$$

where θ is the angle of incidence For perpendicular incidence, $B = 1$. If $V_1 > V$, there is no phase difference between the incident and reflected waves; if $V_1 < V$, there is an 180° phase difference.

Surface reflection. This is essentially perfect at all angles because $\rho_w V_w = 3310 \rho_a V_a$ (the subscripts referring to water and air). However, if the wavelength is small compared to the surface roughness, there is ... instead, the ... the phase ... much of the ... angle of reflection ...

When a sound source is located close to the ocean surface, *reflection* gives rise to surface-image interference or *the Lloyd* mirror effect. Figure 3a indicates how this occurs: Sound from a source S located a distance d_1 below the surface can reach a receiver p at a depth d_2 by two paths, one direct and the other by a reflection from the sea surface (which is here considered perfectly

record of temperature versus depth. This then can be transformed into a record of sound velocity versus depth. *See* BATHYTHERMOGRAPH.

The vertical variation of sound velocity in the deep ocean has three main parts: (1) a variable layer in the first few hundred feet below the surface; (2) a layer, to about 4000 ft, in which the velocity decreases with depth as the temperature decreases in the so-called thermocline; (3) the deepest layer, extending to the ocean floor, in which the velocity increases with depth because of the pressure effect (the temperature being constant at about 37°F). Figure 1 shows a typical velocity-depth dependence that is encountered in the North Atlantic. *See* SEA WATER; THERMOCLINE.

Absorption of sound. Sound is absorbed in sea water because of the irreversible production of heat during the alternate compressions and rarefactions accompanying the sound wave. The absorption is greater in sea water than in pure water because of the effects of dissolved salts, magnesium sulfate having the greatest effect.

Absorption causes a decay in the intensity of a wave which is exponential with distance, that is, the intensity is proportional to $e^{-\alpha R}$, where α is the intensity absorption coefficient and R is the distance traveled. The absorption coefficient α is strongly frequency dependent, being given approximately by the equation

$$\alpha = \frac{40f^2}{4100 + f^2} + 0.000275f^2$$

In this equation, f is the frequency measured in kilocycles per second and α is the absorption coefficient measured in decibels per 1000 yards. This equation holds for a temperature around 50°F, the absorption decreasing somewhat as the temperature increases, the first term is due to the effects of dissolved salts, and the second term is due to viscosity.

Practical cases arise where the measured absorption coefficient can be considerably larger than the value given by the equation; for example, when there is undissolved or entrapped air present, as is the case in wakes of ships, in turbulent or cavitating water, or after a violent storm.

Refraction of sound. This is the bending of sound rays caused by the variation of the sound velocity with ocean depth. Refraction effects are very important when horizontal propagation distances exceed a few hundred feet. Important consequences of refraction to be discussed here are shadow zones and sound channels.

Refraction is described in terms of rays which trace the direction of propagation of the energy. Wavefronts are perpendicular to these rays, and the intensity of sound is proportional to the density of rays. As in geometrical optics, a ray is defined by Snell's law, which can be stated as follows. Let R be the horizontal range and y the depth (measured positively downward), and let it be assumed that the sound velocity V depends only on y, that is, $V = V(y)$. Snell's law then gives $(\cos\theta)/V(y)$ = constant for all the points along a ray, where θ is the angle between the ray and the horizontal direction (θ is taken as positive if downward). If a ray started at an angle θ_0 from a sound source located at a depth $y = 0$, then $(\cos\theta)/V(y) = (\cos\theta_0)/V_0$. The curvature $d\theta/ds$ of a ray at any point (which is the angle the tangent to the curve turns through in a unit distance along the path) is given by

$$\frac{d\theta}{ds} = \left(-\frac{dv}{dy}\right)\left(\frac{\cos\theta_0}{V_0}\right)$$

the variable s being the distance along the ray.

Thus, if the velocity gradient dv/dy is negative, the curvature is positive. This means the angle of inclination increases for distances farther along the ray, which is, therefore, bent in a downward direction. This condition is called downward refraction. The opposite case of upward refraction is found when the velocity increases with depth; that is, the velocity gradient is positive. An important special case is when the velocity-depth dependence is linear: $V(y) = V_0 + Ay$. Then $dv/dy = A$ and the curvature of a ray is constant. Thus, a ray is an arc of a circle of radius

$$r = |V_0/A\cos\theta_0|$$

The centers of the circles which form the rays lie on a line at a depth $y = -V_0/A$.

More complicated situations where the velocity depth dependence is not linear can be approximated by breaking up the velocity curve into small linear segments. Ray-plotting machines also have been devised to draw rays from a knowledge of $V(y)$. The decrease of intensity with distance can be calculated from the spreading of the rays.

In terms of three very common sound velocity depth dependences, refraction has the following effects:

Isothermal surface layer. If there is a superficial layer of constant temperature (called a mixed layer), the velocity increases with depth in the layer because of increasing pressure. For a sound source in the layer, upward refraction occurs for a range of rays above a certain limiting angle $\theta_0{}^1$, these rays reflecting off the surface and following a cyclic path with alternate reflections and refractions (see Fig. 2*a*). Rays which leave the source at angles greater than $\theta_0{}^1$ escape the mixed layer and are refracted downward by the thermocline. Thus, a region beneath the mixed layer is formed where no rays penetrate, called a shadow zone. The region in the mixed layer is called a surface sound channel because sound in it is trapped, and propagation is relatively good. In naval warfare, submarines may escape detection by surface ships by remaining in the shadow zone beneath the mixed layer.

Negative surface gradients. These cause all rays to suffer downward refraction. A shadow zone is formed by that ray which grazes the surface, the so-called limiting ray (see Fig. 2*b*). Propagation is very poor to any point in the [illegible] and the

their presence is observed as scattering layers These layers often have regular daily vertical movements indicating their biological origin. *See* SCATTERING LAYER

Underwater noises. These are caused by marine life, by shipping, by the breaking of waves on beaches, and by such other natural phenomena as storms and rain falling on the surface Thus, at any point in the ocean, there are always present sounds from these and other like causes, not all of which can be identified.

Such sound forms a continuous spectrum of frequencies, being greater in intensity at low frequencies than at high frequencies. When this background noise has no clearly identifiable source, such as nearby marine life or ships, it is referred to as ambient sea noise. This is the result of noise coming from far distant ships and storms, as well as from noise generated at the sea surface. It is found to increase with increasing sea state, indicating that the surface is an important source of noise This ambient noise is the ultimate limiting factor in the detection of far distant sounds. (The term sea state refers to the conditions of the ocean surface; for details, *see* OCEAN WAVES)

Sounds from marine life. These often interfere with the operation of underwater sound detectors (hydrophones) A wide variety of fish produce noises, commonly by the resonant excitation of an air bladder or, occasionally, by the stridulation of teeth The sea robin, for example, drums on the side of an air bladder with

... may produce a loud throbbing roar, far in excess of ambient sea noise Snapping shrimp are another variety of sea creature that is known to produce noise of interfering proportions

Sound is used occasionally by marine creatures themselves for navigation (by echo sounding) and communication purposes as, for example, in the emission of a characteristic distress sound The porpoise, in particular, emits many types of purposeful sounds over a very wide frequency range.

Ship and submarine noise The character of the acoustic energy radiated into the water by a surface ship or submarine varies widely, depending upon the size of the ship, its propulsion machinery, propeller arrangement, and the speed at which it is operating Ship sounds are so characteristic of the type of ship that a trained sonar operator uses them for identification. A ship's spectrum of noise is made up of both a continuous band of frequencies from cavitation and discrete frequencies from the rotation of the propeller and other machinery *Cavitation* is caused by vortices at the propeller blade tips and by water flowing by the ship's hull Cavitation noise also is modulated by the propeller's rotation. Thus there is considerable radiated sound at frequencies corresponding to the propeller blade and shaft frequencies. Submarines emit sounds in much the same way as surface ships, except that there are two effects caused by the depth of submergence. The submarine, being deeper in the water, is a more efficient radiator than a surface ship; also, the onset of cavitation is inhibited by the increased static pressure. *See* CAVITATION; *see also* ECHO SOUNDER; SONAR; SOUND; TRANSDUCER, UNDERWATER; UNDERWATER TELEPHONE. [R.W.MO.]

Bibliography: C. B. Officer, *Introduction to the Theory of Sound Transmission*, 1958.

Underwater telephone

A method of voice communication using underwater sound as the means of transmission. It functions similarly to a conventional telephone system except that the energy is carried by sound waves in the water rather than by electrical signals through a wire. The sound is generated by a *projector* at the sending point and is received by a hydrophone. (For a discussion of projectors and hydrophones, *see* TRANSDUCER, UNDERWATER). Generally, a carrier wave of a fixed frequency is amplitude-modulated by the voice signal and demodulated in the receiving system The range of an underwater telephone is fairly short (perhaps a few miles) because of transmission losses in the sound propagation Its advantage is that it allows communication, without a wire link, beneath the ocean surface where radio wave propagation is impossible. Thus, its most important use is in communication with, or between, fully submerged submarines. *See* UNDERWATER SOUND. [R W MO]

Underwater television

The general principles and techniques employed in underwater television are similar to those used in remotely controlled underwater photography. Both methods have many advantages in common and both are subject to the following limitations: the material or object being studied is only seen and is not brought up for detailed ...

qui

wor

the ... On the other hand, television is the only technique whereby underwater observations can be made continuously at any depth without risk to personnel See TELEVISION; UNDERWATER PHOTOGRAPHY.

Underwater television requires a compact and easily manageable unit that is housed in a watertight casing The casing must be sufficiently strong to withstand the pressure at the greatest workable depth The power supply is fed to the camera and the signals returned from the camera via the camera cable With increasing depth the technical problems of handling the heavy cable become a limiting factor in the use of underwater television, and for work at depths beyond that at which natural light is adequate, a continuous source of artificial light must be provided Usually ordinary tungsten filament lamps are used Monitoring and power units are housed on board ship and the regenerated picture is displayed on one or more

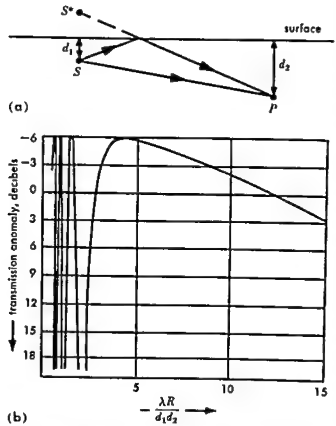


Fig. 3. (a) Lloyd mirror effect for a source close to a calm sea surface. (b) Transmission anomaly versus the parameter $\lambda R/d_1d_2$. Rapid fluctuations occur near the source.

smooth). The reflected sound appears as though it came from a source at S^* which is the image of S. Sound received at P, therefore, is the resultant of two signals, one from S and one from S^*. Depending upon the phase relations between these two sounds, the resultant amplitude at P may be greater or less than the sound amplitude due to S alone. At the surface, the sound amplitude from S^* is equal to, but 180° out of phase with, that from S. Because of this phase reversal, if either S or P is very close to the surface (compared to a wavelength λ) the resultant amplitude, and hence the sound intensity, is nearly zero. It can be shown that the transmission anomaly is given by

$$A = -20 \log \left(2 \sin \frac{2\pi d_1 d_2}{R\lambda} \right)$$

a graph of which is shown in Fig 3b. A becomes infinite (the sound intensity becomes zero) whenever $2d_1d_2/R\lambda = 1, 2, 3, \ldots$ Surface image interference is modified by the roughness of the sea; it is most important when the sea is calm or when the frequency is low.

Bottom reflection. This is more variable than reflection from the surface because properties of the bottom change considerably from place to place. Generally, the sound velocity in the bottom (V_b) is greater than in water, but with muddy bottoms, such as are found in river estuaries, the opposite may occur. Since the density of the bottom is greater than water, the specific acoustic impedance ranges anywhere from 0.8 to $2.5\rho_w V_w$. When $V_b > V_w$, total reflection takes place at angles of incidence greater than the critical angle θ_c given by $\sin \theta_c = V_w/V_b$. Thus, no acoustic energy penetrates the bottom for θ between θ_c and 90°. When $V_b < V_w$, some energy is lost to the bottom at all angles.

Propagation into the bottom will occur as long as $r < 1$. Since the sound velocity in the bottom increases with depth, some of the sound may be refracted back into the ocean. Reflection also can take place from consolidated layers beneath the sedimentary layer. Geophysicists and oceanographers use explosive charges set off in the water for the exploration of these deeper layers.

Scattering of sound. This occurs when a sound wave is incident on air bubbles, marine life, or other objects in the sea. Scattering is that process by which sound energy is redirected from objects (or parts of objects) which have dimensions smaller than a wavelength of sound λ. If the object, or the surfaces of the object, have dimensions and radii of curvatures large compared to λ, the redirection of sound obeys the laws of reflection, and the sound is said to be reflected. It is not always possible to make a clear distinction between reflected and scattered sound.

The simplest cases of scattering for which theoretical solutions can be given are for spheres. If a plane sound wave of intensity I_i is incident on a sphere of radius a, there will be sound scattered in all directions, the scattering pattern being symmetrical about the direction of the incident wave. If ϕ measures the angle of scattering with respect to the incident direction, then the scattered intensity is $I_s(\phi)$, this being always proportional to I_i and, in general, dependent upon λ/a, and the density and compressibility of the material making up the sphere. At distances sufficiently far from the scattering object, $I_s(\phi)$ varies as the inverse square of the distance. If $a \ll \lambda$, $I_s(\phi)$ is found to be proportional to the fourth power of the sound frequency, a case known as Rayleigh scattering.

For a rigid object, not necessarily spherical but much smaller in size than λ, $I_s(\phi)$ is independent of ϕ and is given by

$$I_s = \frac{4\pi^2 B I_i}{\lambda^4 r^2}$$

where r is the distance away from the scattering object, and B is the volume of the object.

Scattering from an air bubble is especially complex because of the high compressibility of air relative to water. Small bubbles can be highly effective sound scatterers because the wave may set them into resonance oscillation.

Scattering from fish may be significant because fish usually possess a gas-filled bladder which is used for buoyancy control. Small marine organisms (such as plankton), when found in large numbers in the sea, also scatter sound measurably. Since these form roughly horizontal layers in the sea,

eir presence is observed as scattering layers. hese layers often have regular daily vertical ovements indicating their biological origin. See CATTERING LAYER.

Underwater noises. These are caused by marine fe, by shipping, by the breaking of waves on eaches, and by such other natural phenomena as torms and rain falling on the surface. Thus, at any oint in the ocean, there are always present sounds rom these and other like causes, not all of which an be identified.

Such sound forms a continuous spectrum of frequencies, being greater in intensity at low frequencies than at high frequencies. When this background noise has no clearly identifiable source, such as nearby marine life or ships, it is referred to as *ambient* sea noise. This is the result of noise coming from far distant ships and storms, as well as from noise generated at the sea surface. It is found to increase with increasing sea state, indicating that the surface is an important source of noise. This ambient noise is the ultimate limiting factor in the detection of far distant sounds. (The term sea state refers to the conditions of the ocean surface; for details, see OCEAN WAVES.)

Sounds from marine life. These often interfere with the operation of underwater sound detectors (hydrophones). A wide variety of fish produce noises, commonly by the resonant excitation of an air bladder or, occasionally, by the stridulation of teeth. The sea robin, for example, drums on the side of an air bladder with special muscles, whereas the trigger fish accomplishes this drumming with fins. Although a single fish does not produce much noise, many of them in concert may produce a loud throbbing roar, far in excess of ambient sea noise. Snapping shrimp are another variety of sea creature that is known to produce noise of interfering proportions.

Sound is used occasionally by marine creatures themselves for navigation (by echo sounding) and communication purposes as, for example, in the emission of a characteristic distress sound. The porpoise, in particular, emits many types of purposeful sounds over a very wide frequency range.

Ship and submarine noise. The character of the acoustic energy radiated into the water by a surface ship or submarine varies widely, depending upon the size of the ship, its propulsion machinery, propeller arrangement, and the speed at which it is operating. Ship sounds are so characteristic of the type of ship that a trained sonar operator uses them for identification. A ship's spectrum of noise is made up of both a continuous band of frequencies from cavitation, and discrete frequencies from the rotation of the propeller and other machinery. Cavitation is caused by vortices at the propeller blade tips and by water flowing by the ship's hull. Cavitation noise also is modulated by the propeller's rotation. Thus, there is considerable radiated sound at frequencies corresponding to the propeller blade and shaft frequencies. Submarines emit sounds in much the same way as surface ships, except that there are two effects caused by the depth of submergence. The submarine, being deeper in the water, is a more efficient radiator than a surface ship; also, the onset of cavitation is inhibited by the increased static pressure. See CAVITATION; see also ECHO SOUNDER; SONAR; SOUND; TRANSDUCER, UNDERWATER; UNDERWATER TELEPHONE. [R.W.MO.]

Bibliography: C. B. Officer, *Introduction to the Theory of Sound Transmission*, 1958.

Underwater telephone

A method of voice communication using underwater sound as the means of transmission. It functions similarly to a conventional telephone system except that the energy is carried by sound waves in the water rather than by electrical signals through a wire. The sound is generated by a *projector* at the sending point and is received by a hydrophone. (For a discussion of *projectors* and hydrophones, see TRANSDUCER, UNDERWATER.) Generally, a carrier wave of a fixed frequency is amplitude-modulated by the voice signal and demodulated in the receiving system. The range of an underwater telephone is fairly short (perhaps a few miles) because of transmission losses in the sound propagation. Its advantage is that it allows communication, without a wire link, beneath the ocean surface where radio wave propagation is impossible. Thus, its most important use is in communication with, or between, fully submerged submarines. See UNDERWATER SOUND. [R.W.MO.]

Underwater television

The general principles and techniques employed in underwater television are similar to those used in remotely controlled underwater photography. Both methods have many advantages in common and both are subject to the following limitations: the material or object being studied is only seen and is not brought up for detailed examination; turbidity quickly reduces picture quality, and, in biological work, the introduction of artificial lights may affect the animals. On the other hand, television is the only technique whereby underwater observations can be made continuously at any depth without risk to personnel. See TELEVISION; UNDERWATER PHOTOGRAPHY.

Underwater television requires a compact and easily manageable unit that is housed in a watertight casing. The [illegible] to withstand [illegible] depth. The pc [illegible] the signals re[illegible] from the camera via the camera cable. With increasing depth the technical problems of handling the heavy cable become a limiting factor in the use of underwater television, and for work at depths beyond that at which natural light is adequate, a continuous source of artificial light must be provided. Usually ordinary tungsten filament lamps are used. Monitoring and power units are housed on board ship and the regenerated picture is displayed on one or more

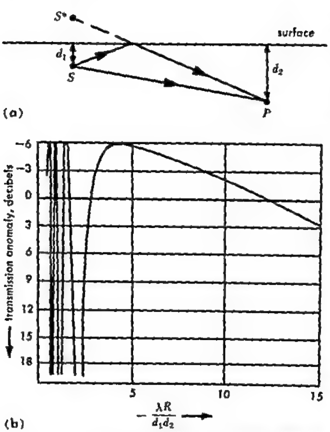


Fig. 3 (a) Lloyd mirror effect for a source close to a calm sea surface. (b) Transmission anomaly versus the parameter $\lambda R/d_1d_2$. Rapid fluctuations occur near the source.

smooth). The reflected sound appears as though it came from a source at S^* which is the image of S. Sound received at P, therefore, is the resultant of two signals, one from S and one from S^*. Depending upon the phase relations between these two sounds, the resultant amplitude at P may be greater or less than the sound amplitude due to S alone. At the surface, the sound amplitude from S^* is equal to, but 180° out of phase with, that from S. Because of this phase reversal, if either S or P is very close to the surface (compared to a wavelength λ) the resultant amplitude, and hence the sound intensity, is nearly zero. It can be shown that the transmission anomaly is given by

$$A = -20 \log \left(2 \sin \frac{2\pi d_1 d_2}{R\lambda} \right)$$

a graph of which is shown in Fig. 3b. A becomes infinite (the sound intensity becomes zero) whenever $2d_1d_2/R\lambda = 1, 2, 3, \ldots$. Surface image interference is modified by the roughness of the sea; it is most important when the sea is calm or when the frequency is low.

Bottom reflection. This is more variable than reflection from the surface because properties of the bottom change considerably from place to place. Generally, the sound velocity in the bottom (V_b) is greater than in water, but with muddy bottoms, such as are found in river estuaries, the opposite may occur. Since the density of the bottom is greater than water, the specific acoustic impedance ranges anywhere from 0.8 to $2.5\rho_w V_w$. When $V_b > V_w$, total reflection takes place at angles of incidence greater than the critical angle θ_c given by $\sin \theta_c = V_w / V_b$. Thus, no acoustic energy penetrates the bottom for θ between θ_c and 90°. When $V_b < V_w$, some energy is lost to the bottom at all angles.

Propagation into the bottom will occur as long as $r < 1$. Since the sound velocity in the bottom increases with depth, some of the sound may be refracted back into the ocean. Reflection also can take place from consolidated layers beneath the sedimentary layer. Geophysicists and oceanographers use explosive charges set off in the water for the exploration of these deeper layers.

Scattering of sound. This occurs when a sound wave is incident on air bubbles, marine life, or other objects in the sea. Scattering is that process by which sound energy is redirected from objects (or parts of objects) which have dimensions smaller than a wavelength of sound λ. If the object, or the surfaces of the object, have dimensions and radii of curvatures large compared to λ, the redirection of sound obeys the laws of reflection, and the sound is said to be reflected. It is not always possible to make a clear distinction between reflected and scattered sound.

The simplest cases of scattering for which theoretical solutions can be given are for spheres. If a plane sound wave of intensity I_i is incident on a sphere of radius a, there will be sound scattered in all directions, the scattering pattern being symmetrical about the direction of the incident wave. If ϕ measures the angle of scattering with respect to the incident direction, then the scattered intensity is $I_s(\phi)$, this being always proportional to I_i and, in general, dependent upon λ/a, and the density and compressibility of the material making up the sphere. At distances sufficiently far from the scattering object, $I_s(\phi)$ varies as the inverse square of the distance. If $a \ll \lambda$, $I_s(\phi)$ is found to be proportional to the fourth power of the sound frequency, a case known as Rayleigh scattering.

For a rigid object, not necessarily spherical but much smaller in size than λ, $I_s(\phi)$ is independent of ϕ and is given by

$$I_s = \frac{4\pi^2 B I_i}{\lambda^4 r^2}$$

where r is the distance away from the scattering object, and B is the volume of the object.

Scattering from an air bubble is especially complex because of the high compressibility of air relative to water. Small bubbles can be highly effective sound scatterers because the wave may set them into resonance oscillation.

Scattering from fish may be significant because fish usually possess a gas-filled bladder which is used for buoyancy control. Small marine organisms (such as plankton), when found in large numbers in the sea, also scatter sound measurably. Since these form roughly horizontal -in the sea

lated to a fundamental tensor g_{km} by certain linear equations. When g_{km} is symmetric, space time may be regarded as metric, and the theory of pure gravitation results. Einstein expected that the skew-symmetric part of g_{km} should be related to electromagnetism. V. Hlavatý proved that it is indeed possible to define in terms of g_{km} tensors which satisfy the classical electromagnetic field equations.

A diametrically opposed approach, and one which is far less speculative, is due to F. Kottler. Recognizing the artificiality of space-time metrics, he sought to show that physical laws are largely independent of geometry. He easily succeeded in formulating electromagnetic theory for any smooth four-dimensional manifold, no geometric structure being assumed. R. Toupin showed that a correspondingly general theory of mechanics (not merely gravitation) is possible by means of tensors depending upon the positions of two observers. These fully invariant formalisms state the known physical laws with a minimum of assumptions, leaving the way open to imposing such restrictions as future physical experience may show to be necessary. Any structure that space-time possesses appears as a result, not as the cause, of physical laws. See SPACE-TIME; see also CALCULUS OF TENSORS; ELECTROMAGNETISM; GRAVITATION; LORENTZ TRANSFORMATIONS; MAXWELL'S EQUATIONS; RELATIVISTIC MECHANICS; RELATIVITY.

Bibliography. A. Einstein, *The Meaning of Relativity*, 5th ed., 1955; E. Schrödinger, *Space-Time Structure*, 1950; E. T. Whittaker, *A History of the Theories of Aether and Electricity*, vol. 2, 1954.

Unit operations

A concept used by chemical engineers to teach and practice industrial processing. The basis of the unit operation approach is that chemical processes, in spite of their great variety, may be resolved into a comparatively small number of units, each of which appears again and again in the various processes. E[illegible] [illegible] ed by a system of vaporizations and condensations. This operation is used in petroleum refining, alcohol manufacture, liquid gas production, refining of solvents, and many other industries. Knowledge of the fundamentals of distillation is applicable, then, to all these industries. This is characteristic of all unit operations.

The unit operations alone are insufficient for treating all processing problems. They [illegible] supplemented [illegible] and scienti[illegible] control, ap[illegible] [illegible] of construction [illegible] chemical engineering, both educationally and in practice, consists of a correlated treatment and application of all this subject matter.

The list of unit operations is neither rigid nor permanent. New ones are added from time to time as they become useful. One list, reflecting current

Unit operations*

Name	Function
Operations based on fluid mechanics	
Fluid handling	Storing and transporting fluids. Controlling and measuring flow of gas, liquid, and vapor
Mixing and agitation	Combining liquids, solids, and gases into homogeneous fluids or intimate mixtures of phases
Filtration and clarification	Separating solid particles from liquids and gases
Thickening and sedimentation	Concentrating solids from their mixtures with liquids
Classification	Sorting solid particles by size and specific gravity
Centrifugation	Separating solids from liquids and liquids from liquids by centrifugal force
Operations based on heat transfer	
Heat exchange and condensation	Heating, cooling, and condensing of fluids with or without phase change
Furnaces and kilns	High temperature heating of materials
Evaporation and boiling	Vaporizing liquids, concentrating solutions of nonvolatile solids, recovering distilled water
Drying	Removing moisture or other liquid from solids by vaporization or other means
Cooling towers	Cooling water for reuse in condensers or for air conditioning
Operations based on mass transfer	
Distillation	Separating miscible liquids by vaporization
Liquid extraction	Separating miscible liquids by differences of solubility
Leaching	Dissolving soluble substances from solids by solvent extraction
Absorption and desorption	Washing soluble gas from its mixture with inert gas by liquid. Removing dissolved gas from liquid by inert gas
Adsorption	Selectively removing substances from liquids or gases by reactive solids
Ion exchange	Selectively exchanging ions in solutions of electrolytes by reactive solids
Humidification and dehumidification	Controlling humidity or vapor content of air or gas
Diffusion of gases	Separating mixtures of gases by temperature gradients or by other specialized methods
Operations based on mechanical principles	
Screening	Separating particles according to size by screens
Solids handling	Transporting and storing solids
Size reduction	Subdividing solids into smaller particles
Flotation	Separating solids by selective aeration
Magnetic and electrostatic separations	Separating, by electrical methods, solids according to chemical composition, or solids from gases

* See separate articles on many of these unit operations.

standard picture monitors. The screen picture may be photographed by either still or moving-picture cameras if a permanent record is required. The image may also be transmitted from ship to shore, but generally the picture is viewed by closed-circuit television aboard the ship. *See* CLOSED-CIRCUIT TELEVISION.

Several types of television camera tubes are used, namely, the cathode-potential-stabilized (c-p-s) emitron, image orthicon, and vidicon The c-p-s emitron, in the opinion of many, gives the best picture quality but lacks the extreme sensitivity of the image orthicon With vidicon both picture quality and sensitivity are less satisfactory; nevertheless, vidicon provides the most compact and easily handled unit and therefore is used for some purposes. For the most detailed work, particularly of a biological character, the remote control of the iris, focus, lens selection, and, if possible, the lighting is desirable. *See* TELEVISION CAMERA TUBE.

Apart from its use by naval personnel, underwater television has been employed in dockyard engineering, in the examination of underwater equipment, and in biological investigations of marine and freshwater habitats It has been used in both qualitative and quantitative ecological studies of bottom fauna, in plankton studies, and in examination of the inanimate environment of animals on the sea bed. *See* OCEANOGRAPHY [H.B.]

Bibliography: R. H Backus and H Barnes, Television-echo sounder observations of midwater sound scatterers, *Deep-Sea Research*, 4(2) .116–119, 1957; H. Barnes, *Oceanography and Marine Biology*, 1959; H. Barnes, Underwater television and research in marine biology, bottom topography and geology, *Deut. Hydrograph Z*, pt. 1, 6·123–133. 1953, pt. 2, 8:213–236, 1955.

Unified field theories

Classical theories that represent gravitation and electromagnetism as aspects of space-time geometry Although other field theories, such as quantum field theory which combines quantum mechanics and relativity theory, may be called unified field theories, the term as commonly used in the literature refers only to the attempts to unify the theories of gravitational and electromagnetic forces.

Historical development. Motivation for the development of unified field theories can scarcely be comprehended except in view of earlier results in special and general relativity. The relativity principle of H. Poincaré and H. A. Lorentz grew in part from the discovery that the classical electromagnetic field equations of J. C. Maxwell preserve their form under certain transformations of space and time known as Lorentz transformations. M. Planck showed how to modify the laws of dynamics so as to render them consistent with the relativity principle. H. Minkowski unified these concepts by picturing mechanics and electromagnetism in a four-dimensional noneuclidean manifold, called space-time, in which any separation into a three-dimensional space accompanied by a one-dimensional time is merely arbitrary and accidental. He found that phenomena in space-time are conveniently and invariantly described by means of the tensor analysis of G. Ricci-Curbastro, and he found the tensorial equation expressing conservation of relativistic energy-momentum.

Meanwhile, A. Einstein had conjectured that gravitation, which was left out of account in special relativity, is in reality only an appearance arising from the motion of the observer. H. Bateman found that the electromagnetic equations are invariant under a much larger group of transformations than those considered in special relativity, and he introduced arbitrary coordinates in a four-dimensional metric space-time, allowing for relative acceleration of observers. This idea was expanded by Einstein, who constructed a curved space-time in which the metric tensor itself determines the paths of particles and thus is a means of specifying the gravitational field.

D Hilbert found a fully invariant single law, which encompassed both Einstein's equations of gravitation and Maxwell's equations of the electromagnetic field. Formally, the two theories are united, but still there is a difference While gravitation is completely identified with the metric properties of space-time, so that the notion of a mechanical force on ponderable bodies due to gravitational attraction is completely abolished, the notion of a mechanical force acting on electrified or magnetized bodies is retained, as in the older physics.

Unification proposals. There is no experimental evidence for a connection between electromagnetism and gravitation The need felt for a unified theory is of philosophic or aesthetic origin. Moreover, unification from a purely mathematical standpoint is easy to achieve in many different ways Thus the subject remains not only speculative but personal A few of the better-known proposals are mentioned here

In the theory of H Weyl, four-dimensional space-time is characterized not only by the quadratic form, $g_{km}\, dx^k\, dx^m$ (where g_{km} is a tensor), of the relativistic theory of gravitation but also by a linear form, $\phi_p\, dx^p$ The coefficients, ϕ_p, of the latter are to be interpreted as a four-dimensional electromagnetic potential. Weyl later retracted this theory

In the theory of T. Kaluza and O. Klein, the potentials g_{km} and ϕ_p are expressed in terms of the metric tensor of a five dimensional flat space time. The space-time of physical experience is considered to approximate a thin four-dimensional curved shell, its fifth dimension being very short, and ordinary measurements are to be interpreted as averages over this fifth dimension.

Einstein himself, in several trials, always adhered to a four-dimensional structure. His last theory regards space time as affine rather than metric, although the affine connection is assumed to be re-

a large number of industries, the unit processes or chemical changes are conducted in equipment made of metals such as iron, cast iron, steel, stainless steel, copper, nickel, and nonferrous alloy such as Hastelloy, but in certain highly corrosive conditions, glass-lined or even pure glass equipment is employed. Occasionally wood or plastic-coated or -impregnated material operates satisfactorily. For example wood is frequently used in tanks and filter presses in the dyestuff industry.

Very few chemical reactions are neutral thermally. Heat is either given out or required. Consequently, the unit operation of heat transfer is intimately connected with the chemical reaction. It is likewise true of mixing, pumping, and other physical changes.

Significance of concept. Because of the greater simplification and the importance of investigations pertaining to unit operations, these physical changes have been studied intensively since 1910 and great progress has been made to the benefit of the chemical industry. However, because there is no more important factor in the reduction of costs in the manufacture of chemicals than that of increasing the yields and conversions, the chemical or unit process aspect is now receiving increasing attention. An over-all term has been evolved in the last few years to embrace the various unit processes: chemical processing. Chemical processing often involves a sequence of the unit processes combined with unit operations and carried out on an industrial scale.

The unit process concept is useful because of, or influenced by, the following factors.

Each unit process emphasizes the unitary or common aspects of numerous individual reactions.

In industry, factors segregation by the unit process leads to a higher time percentage of utilization for smaller scale production and to greater availability of experienced personnel for all scale productions.

The concept permits multiple use of equipment for many individual chemicals.

The unit process classification emphasizes principles and this facilitates design as well as production. Both the chemical engineer in charge and the workmen can utilize their experience in a more favorable environment. They become specialists in hydrogenations, nitrations, or other processes.

Because the basis of the unit-process concept is the chemical reaction, equilibrium, kinetics, and catalysis are emphasized. Increases in the chemical conversion or yield are by and large the most important factors in reducing costs of chemicals. The cost of raw materials represents 50–80% of the manufacturing expense for most chemicals.

No distinction need be made between inorganic and organic procedures because remarkably similar conditions in a given unit process prevail for both, for example, in the hydrogenation of N_2 to NH_3 or of CO to CH_3OH.

In design of equipment, the engineer is greatly aided by the generalizations arising from the unit-process arrangement rather than by considerations of reactions separately.

For additional details, see the separate articles on the individual unit processes. *See also* CHEMICAL ENGINEERING; UNIT OPERATIONS. [R.N.S.]

Bibliography: R. N. Shreve, *Chemical Process Industries*, 2d ed., 1956.

Units, systems of

Groups of units suitable for use in the measurement of physical quantities and in the convenient statement of physical laws relating physical quantities. Quantities are the quantitative concepts of physical science. A given physical quantity A such as length, time, or energy, is the product of a numerical value or "measure" $\{A\}$ and a unit $[A]$, the unit being a particular reference-sample of the quantity chosen as a standard. Thus,

$$A = \{A\}[A] \qquad (1)$$
$$\text{Quantity} = \text{numerical value} \times \text{unit}$$

The unit $[A]$ can be chosen quite arbitrarily, but it is desirable to define units in such a way that they are derived from a few fundamental units by equations without numerical factors other than unity, and that the equations between numerical values of quantities have exactly the same form as the equations between the quantities. For example, the kinetic energy E of a body is given by

$$E = \tfrac{1}{2}MV^2 \qquad (2)$$

where $E = \{E\}[E]$, $M = \{M\}[M]$, $V = \{V\}[V]$, and $\tfrac{1}{2}$ is called a definitional factor and is dimensionless. If the units of E, M, and V are defined in such a way that

$$[E] = [M][V]^2 \qquad (3)$$

then the equation between the numerical values is

$$\{E\} = \tfrac{1}{2}\{M\}\{V\}^2$$

A system of units defined in this way is called a coherent system. It is constructed by defining the units of a few fundamental quantities independently; these are called fundamental units. The units of all other quantities are defined by equations similar to Eq. (3) with no numerical factors other than unity and are called derived units.

Absolute systems. It is a matter of choice which quantities are to be considered as fundamental and also how many. It is desirable to have as few fundamental quantities as possible, provided no serious inconveniences are encountered. In mechanics, it is convenient to have three fundamental quantities. In the so-called absolute systems these quantities are length, mass, and time. Two absolute systems of metric units are commonly employed: (1) the mks (meter-kilogram-second) system which has been adopted by international commissions, and (2) the cgs (centimeter-gram-second) absolute system. Once the length, mass, and time units are selected, Newton's second law in the form $F = ma$ is used to define the force unit in terms of mass units and acceleration units, which involve

usage, is given in the table. Each operation is listed, and its function summarized.

The unit operations approach to processing originated early in the present century and became popular during the decade 1920–1930. At first, the individual operations were considered to be quite distinct and unrelated. Active study and research in this field since 1930 have achieved considerable unification within the unit operations area. It has been shown that the unit operations are themselves based on an even smaller group of scientific areas. Thus, fluid mechanics, heat transfer, and mass transfer provide the foundations for most of the individual operations, and these three engineering sciences are themselves closely related. The foundation for the remaining unit operations is mechanics. The operations listed in the table are classified under the appropriate foundation sciences. Recent trends in teaching the unit operations emphasize the sciences of fluid mechanics, heat transfer, and mass transfer, and then apply these subjects to the separate operations.

The theoretical foundations of the unit operations are useful only because they lead to sound designs of the equipment for using them in an actual plant. There is greater variety in the equipment than in the theoretical laws. Pumps, piping systems, dryers, heat transfer equipment, distilling columns, crushers, mixers, absorbers, evaporators, conveyors, and centrifuges are typical classes of unit operations equipment. Each kind is built in a wide variety of types and designs. The final objective of the unit operation approach is to provide the plant designer with sound selections and designs of practical apparatus for efficient production. See CHEMICAL ENGINEERING [W.L.M.]

Bibliography: W. L. McCabe and J. C. Smith, *Unit Operations of Chemical Engineering*, 1956.

Unit processes

Unit processes have also been named unit chemical changes or conversion processes. Examples of such changes are oxidation, hydrogenation, sulfonation, nitration, diazotization, and about 20 others. These all refer to the unit chemical changes into which some chemical processes can be divided. Unit processes have been defined by R. N. Shreve as "the commercialization of a chemical reaction under such conditions as to be economically profitable." To supplement these, there are unit operations or unit physical changes. These two make up the units into which a chemical process, as represented by a flow sheet, can be broken down to depict graphically the complete process in proceeding from raw material to finished product in the chemical industry.

Such fundamental success was experienced in the early days of chemical engineering in breaking physical changes into various unit operations that a similar classification was started about 1930 for the chemical changes. However, the chemical changes are much more complicated than the physical ones, and, hence, chemical engineers have not succeeded in classifying each individual chemical change as scientifically as has been done for the unit operations or physical changes. The principal unit processes are

Alkylation	Hydrogenation, hydrogenolysis
Ammonolysis	Hydrolysis, hydration
Aromatization	Ion exchange
Calcination, dehydration	Isomerization
Combustion	Neutralization
Condensation	Nitration
Diazotization and coupling	Oxidation
Electrolysis	Polymerization
Esterification (sulfation)	Pyrolysis, cracking
Fermentation	Reduction
Halogenation	Sulfonation
Hydroformylation (oxo)	

Basic principles. Certain unifying aspects of these chemical changes exist simply because they all embody chemical reactions. The most important from an industrial or chemical engineering viewpoint involves the laws for chemical equilibria and kinetics. Although both the chemical equilibria and kinetics vary greatly from one unit process to another, the general principles are the same for all processes.

In many cases, catalysts can be employed to speed up the reaction in order to reach the equilibrium within a practical time. Another common factor among almost all unit processes, is the importance of conversion and equilibrium, the conversion being expressed usually as molar per cent of the main raw material that is changed chemically into the main product in a single pass through the system. The equilibrium is expressed by the percentage net molecular change. This is determined by subtracting the moles recovered of the principal raw material from the moles chemically charged, and equating chemically this net with the moles formed of the main product. The aim is to raise the chemical conversion to the chemical equilibrium and to do this within a practical time. Chemical equilibria can be raised in specific gases by changing the proportions of reactants and by altering time, temperature, and pressure. The following reaction illustrates yield and conversion:

$$N_2 + 3H_2 \rightleftharpoons 2NH_3$$

At about 5000 lb pressure and 500°C, the yield will be 99+% but the conversion will be only about 20%. In other words, only 20% of the moles charged will be converted into ammonia per pass, but in this case, there are practically no losses or by-products, so the yield or net change is high. On the other hand, in many reactions, there are many by-products which cut down the yields and conversions. This is true in many oxidations and in many nitrations, especially in processes using the aliphatic series of hydrocarbons. See NITRATION.

The equipment needed to carry out reactions is of vital significance in any a[illegible]ons. In

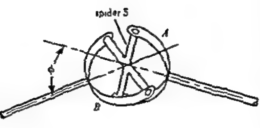


Fig 1 Hooke's joint.

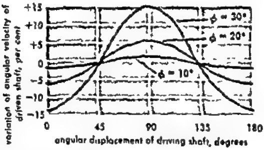


Fig 2. Variation of angular velocity of driven shaft of Hooke's joint for constant angular velocity of driving shaft

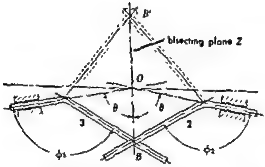


Fig 3 Intersecting shafts in sliding contact maintain constant velocity *(R. T. Hinkle, Kinematics of Machines, Prentice-Hall, 1953)*

shaft B is the same as that of shaft A only at the end of each 90° of shaft rotation Thus, the angular velocity of shaft B, ω_B, is the same as the angular velocity of shaft A, ω_A, only at four positions during each revolution Three curves, showing deviation of ω_B from a constant ω_A as shaft A is turned through 180°, are plotted in Fig. 2 The figures on the curves represent ϕ, the angle between shafts

Double Hooke's joint. The variation in angular displacement and angular velocity between driving and driven shafts, which is objectionable in many mechanisms, can be eliminated by using two Hooke's joints, with an intermediate shaft. This arrangement is conventional for an automobile drive shaft The axes of the driving and driven shafts need not intersect, but it is necessary that the axes y and y' of the two yokes attached to the intermediate shaft lie respectively in planes containing axes of adjoining shafts (a, y, and b in same plane; b, y', and c in same plane), and that the angle between the driving and intermediate shafts equal the angle between the intermediate and driven shafts.

Bendix-Weiss joint. Two intersecting thin shafts (Fig 3) will maintain constant angular velocity ratio, and the point of contact between the shafts will always lie in plane Z, which bisects the angle between shafts and is perpendicular to the plane containing the axes of the shafts

One practical application of this principle, in which constant angular velocity is transmitted through a single universal joint, is the joint developed by Carl Weiss and illustrated in Fig. 4 The four large balls are transmitting elements, while the smaller center ball acts as a spacer. The transmitting balls must lie in the plane Z, as explained above By means of milled grooves in the yoke

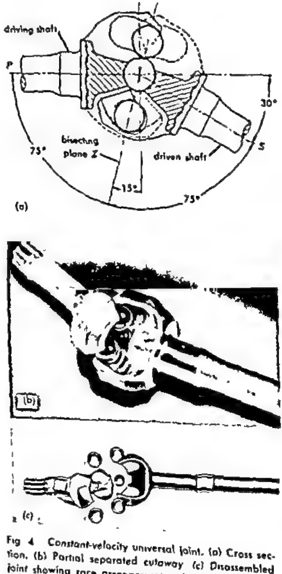


Fig 4 Constant-velocity universal joint. (a) Cross section, (b) Partial separated cutaway (c) Disassembled joint showing race arrangement with right-hand member cut away (Bendix Aviation Corp.)

Table 1. Absolute systems of units

Quantity	Defining relation	Units: mks	Units: cgs	Units: British
Length, L	Fundamental	m	cm	ft
Mass, m	Fundamental	kg	g	pound mass (lbm)
Time, t	Fundamental	sec	sec	sec
Velocity	$v = L/t$	m/sec	cm/sec	ft/sec
Acceleration	$a = v/t$	m/sec²	cm/sec²	ft/sec²
Force	$F = ma$	1 newton = 1 kg-m/sec²	1 dyne = 1 g-cm/sec²	1 poundal = 1 lbm-ft/sec²
Work	$W = FL$	1 joule = 1 newton-m	1 erg = 1 dyne-cm	poundal-ft
Power	$P = W/t$	1 watt = 1 joule/sec	erg/sec	ft-poundal/sec
Momentum	$p = mv$	kg-m/sec	g-cm/sec	lbm-ft/sec

Table 2. Gravitational systems of units

Quantity	Defining relation	Units: cgs	Units: British
Length, L	Fundamental	cm	ft
Force, F	Fundamental	gram-force (gf)	pound force (lbf)
Time, t	Fundamental	sec	sec
Mass	$m = F/a$	gf-sec²/cm	slug
Velocity	$v = L/t$	cm/sec	ft/sec
Acceleration	$a = v/t$	cm/sec²	ft/sec²
Work	$W = FL$	cm-gf	foot-pound force (ft-lbf)
Power	$P = W/t$	cm-gf/sec	ft-lbf/sec
Momentum	$p = mv$	gf-sec	slug-ft/sec

length and time units. The unit of force derived in this way through Newton's second law is called the newton in the mks system and the dyne in the cgs system. Work units are defined by the relation $W = FL$, where L represents a displacement in the direction of a force F. Power units in turn are defined by the relation $P = W/t$ and are therefore the ratio of work units to time units. A coherent absolute system of British units is based on the foot, the pound mass [1 pound mass (lbm) = 0.4536 kg], and the second, and the derived unit of force is called the poundal, a unit which is not often used. Various quantities expressed in three absolute systems are listed in Table 1.

Gravitational systems. Another type of system frequently employed by engineers is a gravitational system in which length, force, and time are regarded as fundamental. In the British gravitational system the standard force, called the pound force (lbf), is the gravitational force exerted on a pound mass at a location on the earth's surface where the acceleration of a freely falling body is 32.17398 ft/sec². A coherent system is obtained by using Newton's second law, $F = ma$, to define a mass unit called the slug (1 lbf = 1 slug-ft/sec²). The work unit in this system is the foot-pound force (ft-lbf).

In the cgs gravitational system the unit of acceleration is the centimeter per second per second and the unit of force is taken to be the earth's gravitational force on a 1-gram mass at a specified location. Unfortunately this unit of force is also called a gram, but to distinguish it from the unit of mass, it is often called gram force. No name is given to the cgs gravitational unit of mass. Various quantities expressed in the cgs gravitational and the British gravitational systems are listed in Table 2.

Noncoherent gravitational systems in which force is expressed in pounds force (lbf) or a comparable kilogram force (kgf) and mass in pounds mass (lbm) or kilograms can be set up by writing Newton's second law in the form $g_0F = ma$, where g_0, a proportionality constant, is numerically equal to a specified value of the acceleration due to gravity. See DIMENSIONAL ANALYSIS; ELECTRICAL UNITS; PHYSICAL MEASUREMENT. [D.W.L.]

Bibliography. W. E. Forsythe (ed.), *Smithsonian Institution Physical Tables*, 9th ed., 1956; Nat. Acad. Sci. Nat. Research Council, *International Critical Tables*, 7 vols., 1926–1930; R. L. Weber, M. W. White, and K. V. Manning, *Physics for Science and Engineering*, 1957.

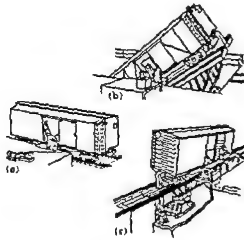
Universal joint

A linkage whose purpose is to cause mutual rotation of two shafts whose axes do not coincide. The joint is used in almost every class of machinery—machine tools, instruments, control devices, and, most familiarly, automobiles.

Hooke's joint. A simple universal joint, known in English-speaking countries as Hooke's joint and in continental Europe as a cardan joint (see FOUR-BAR LINKAGE), is shown schematically in Fig. 1. It consists of two yokes A and B, attached to their respective shafts and connected by means of spider S. Angle ϕ between shafts may have any value up to approximately 35°, if the angular velocity is moderate when the angle ϕ is large. While shaft B must make one revolution for each revolution of shaft A, the instantaneous ar[illegible] [illegible]lacement[illegible]

inloader

'ower device for removing bulk materials from beeath or within motor trucks and rail cars. Forms f portable belt, drag chain, flight and screw coneyors handle freely flowing materials discharged s self-dumping carriers. Machines, such as belt conveyors, power shovels, scoops, and pneumatic



Unloader. (a) Power scoop (b) Rocking type (c) Tilting type (Link-Belt Co.)

devices, are adapted to unloading nondumping trucks and cars. A mechanized unloader includes a section of track which can be rocked to unload open top (gondola) cars and, if fully mechanized, it can be tilted to unload boxcars. Unloading rates up to 10 boxcars per hour are achieved in this way. See BULK HANDLING MACHINES [D O H]

Unsaturated hydrocarbon

One of a class of hydrocarbons that have at least one double or triple carbon-to-carbon bond that is not in an aromatic ring. Hydrocarbons that can undergo addition reactions involving a pair of adjacent atoms, or atoms which are conjugated, are said to be olefinic or unsaturated. With the limitations defined above, addition reactions involving ring opening such as the addition of halogen acid to cyclopropane or to an oxirane, or additions to a single atom which undergoes a change in valence, are not used as criteria of unsaturation. Unsaturated hydrocarbons may be acyclic (open-chain) or alicyclic (closed-chain or ring) compounds. Unsaturated compounds may be classified as alkenes, characterized by the presence of one double bond, $-C{=}C-$, polyalkenes, which have two or more unsaturated linkages, or alkynes, which have triple bonds. See ALKENE; ALKYNE; POLYALKENE.

Olefinic hydrocarbons are named under the rules of the International Union of Pure and Applied Chemistry, more commonly known as the Geneva nomenclature. See ORGANIC CHEMISTRY.

Physical properties. Olefins may be gases, liquids, or crystalline or waxy solids. Most of the theoretically possible lower members of the homologous series of olefinic hydrocarbons have been synthesized in a pure state, but the astronomically large number of isomers which are possible in compounds of higher molecular weight has made the task of preparing a complete series of olefinic hydrocarbons exceedingly difficult. In addition to structural isomerism, olefins with nonterminal double bonds may exist as stereoisomers and are present in conventionally synthesized products as mixtures of the *cis* and *trans* forms (see ISOMERISM, MOLECULAR). Some of the more important unsaturated hydrocarbons and their physical constants are listed on the next page.

Occurrence and preparation. Commercially important unsaturated hydrocarbons such as ethylene, propylene, the butenes, butadiene, isoprene, cyclopentadiene, and acetylene are produced on a tonnage basis by the petroleum industry by means of thermal or catalytic cracking processes. In the

out catalysts such as active clays, acids, or iodine; pyrolysis of esters of alcohols or glycols; and thermal decomposition of xanthate esters (Tschugaeff synthesis), quaternary bases (Willstätter synthesis), and trialkyl amine oxides. Methods which have been especially useful for the synthesis of alpha olefins are the Grignard synthesis and Boord synthesis. The Boord synthesis employs reduction of alpha bromo ethers with zinc. Since the two stereoisomeric forms of a particular olefin usually have very close boiling points, it is difficult to separate them by distillation, and special methods have been developed for their synthesis. These include simultaneous dehydrohalogenation and decarboxylation of halogen substituted acids and partial reduction of substituted alkynes.

Chemical properties. Unsaturated hydrocarbons may undergo polymerization, cyclization, and addition reactions. A major share of structural and elastic polymers are based on homopolymers or copolymers of olefins and diolefins. Thus, polyethylene is made today by means of high-pressure peroxide-catalyzed processes or by low pressure methods using Ziegler-type catalysts. Similar to polyethylene is polypropylene made by Ziegler-type catalysts. Polystyrene is made by emulsion polymerization of styrene, and a similar process is used for the production of GR-S synthetic rubber from styrene and butadiene. Butyl rubber from isobutylene and butadiene or isoprene is made by means of aluminum chloride at low temperature. Polybutenes made by aluminum chloride or boron fluoride polymerization of isobutylene or isobutylene-butene mixtures vary in consistency from oils to rubbery polymers. An exact duplicate of the stereochemical structure of polyisoprene in natural rubber has recently been made by using lithium catalysts. Synthetic drying oils have been made from butadiene and butadiene-styrene copolym

attached to each shaft, the balls are maintained at all times in plane Z. This joint, used for a front-wheel automotive drive, transmits unvarying angular displacement even when steering of the vehicle requires a varying angle between driving and driven shafts. [E.S.F.]

Universal motor

A series motor built to operate on either alternating current (ac) or direct current (dc). It is normally designed for capacities less than one horsepower. It is usually operated at high speed, 3500 rpm loaded and 8000 to 10,000 rpm unloaded. For lower speeds, reduction gears are often employed. Like all series motors, the speed increases as the load decreases and the no-load speed is limited only by friction and windage. To obtain more constant speed, a centrifugal governor may be used to switch in or out a small resistor in series with the armature as in Fig. 1.

If an alternating current is applied to any dc series motor, the motor would still rotate. Since the current is reversed simultaneously in the armature and the field, the torque would pulsate but would not reverse direction. However, a universal motor designed to operate on ac should have laminated cores to avoid excessive eddy currents. Also, there are usually fewer turns in the field coils than in a dc motor. *See* CORE LOSS; DIRECT-CURRENT MOTOR.

The series ac motor is an alternating-current commutator motor which has great flexibility of performance. It can be operated over a wide range of speeds and is readily controllable. The series ac commutator motor is in many respects similar to the dc series motor and the universal motor.

The ac series motor, like the dc series motor and the repulsion motor, consists fundamentally of these windings or their equivalent: (1) rotating armature winding, (2) stationary field winding, and (3) compensating winding. See Fig. 2.

A major problem in larger ac series motors is in commutation. Because of the transformer action between the field and armature coils, voltage is produced in the armature coils which are short-circuited by the brushes as the commutator bars pass under them. The coils which are short-circuited act like a short-circuited secondary of a static transformer. The resulting large currents are interrupted as the bars pass the brushes, causing bad sparking. In addition, these induced currents reduce the magnetic flux of the field and reduce the torque of the motor. Interpoles shunted with non-

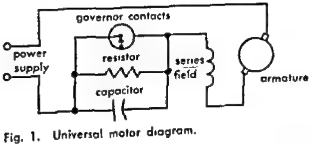


Fig. 1. Universal motor diagram.

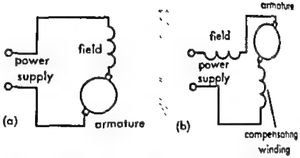


Fig. 2 Series motor diagram. (a) Without compensating winding. (b) With compensating winding.

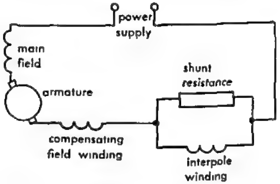


Fig. 3 Single-phase traction motor circuit diagram (*General Electric Co.*)

inductive resistance are required on ac series motors as in Fig. 3. *See* COMMUTATION.

The single-phase commutator motor usually has a large number of turns in the armature winding and a small number of turns in the field winding, as compared with the dc motor which is designed for relatively strong field and weaker armature. The ac motor usually has a larger number of poles and operates at a lower voltage than its dc counterpart. [S.W.]

Bibliography: A. E. Knowlton (ed.), *Standard Handbook for Electrical Engineers*, 9th ed., 1957; R. R. Lawrence and H. E. Richards, *Principles of Alternating-current Machinery*, 4th ed., 1953.

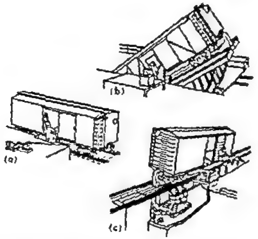
Universe

The totality of things, events, relations, and energies that are capable of being described objectively. The universe contains, or possibly coincides with, the physical universe, which in turn contains the observable universe. The term observable universe means either the set of presently observable phenomena or the set of phenomena that could be observed with ideal instruments, such as telescopes of infinite light-gathering ability. Model universes form the subject of cosmology.

Theories usually deal with the observable universe because it has a definite boundary; the physical universe may be assumed to be unbounded. Because theories of the physical universe make predictions about the observable universe, they are in principle as susceptible to observational verification as any other physical theory. *See* COSMOLOGY; GALAXY, EXTERNAL; METAGALAXY.

Unloader

Power device for removing bulk materials from beneath or within motor trucks and rail cars. Forms of portable belt, drag chain, flight and screw conveyors handle freely flowing materials discharged by self-dumping carriers. Machines, such as belt conveyors, power shovels, scoops, and pneumatic



Unloader (a) Power scoop (b) Rocking type (c) Tilting type. (Link-Belt Co)

devices, are adapted to unloading nondumping trucks and cars A mechanized unloader includes a section of track which can be rocked to unload open top (gondola) cars and, if fully mechanized

Unsaturated hydrocarbon

One of a class of hydrocarbons that have at least one double or triple carbon-to-carbon bond that is *not in* an aromatic ring. Hydrocarbons that can undergo addition reactions involving a pair of adjacent atoms, or atoms which are conjugated, are said to *be olefinic* or unsaturated With the limitations defined above, addition reactions involving ring opening such as the addition of halogen acid to cyclopropane *or to* an oxirane, or additions to a single atom which undergoes a change in valence, are not used as criteria of unsaturation Unsaturated hydrocarbons *may be acyclic* (open-chain) or alicyclic (closed-chain, or ring) compounds Unsaturated compounds *may* be classified as alkenes, characterized by the presence of one double *bond*, —C=C—, polyalkenes, which have two or more unsaturated linkages, or alkynes, which have triple bonds See ALKENE, ALKYNE, POLYALKENE.

Olefinic hydrocarbons are named under the rules of the International Union of Pure and Applied Chemistry, more commonly known as the Geneva nomenclature *See* ORGANIC CHEMISTRY

Physical properties. Olefins may be gases, liquids, or crystalline or waxy solids Most of the theoretically possible lower members of the homologous series of olefinic hydrocarbons have been synthesized in a pure state, but the astronomically large number of isomers which are possible *in* compounds of higher molecular weight has made the task of preparing a complete series of olefinic hydrocarbons exceedingly difficult. In addition to structural isomerism, olefins with nonterminal double bonds may exist as stereoisomers and are present in conventionally synthesized products as mixtures of the *cis* and *trans* forms (see ISOMERISM, MOLECULAR). Some of the more important unsaturated hydrocarbons and their physical constants are listed on the next page.

Occurrence and preparation. Commercially important unsaturated hydrocarbons such as ethylene, propylene, the butenes, butadiene, isoprene, cyclopentadiene, and acetylene are produced on a tonnage basis by the petroleum industry by means of thermal or catalytic cracking processes. In the laboratory, most methods for the synthesis of hydrocarbons involve elimination reactions: CX—CY → C=C + X—Y Methods which have been used include dehydration of alcohols either with or without catalysts such as active clays, acids, or iodine; pyrolysis of esters of alcohols or glycols; and thermal decomposition of xanthate esters (Tschugaeff synthesis), quaternary bases (Willstätter synthesis), and trialkyl amine oxides. Methods which have been especially useful for the synthesis of alpha olefins are the Grignard synthesis and Boord synthesis The Boord synthesis employs reduction of alpha bromo ethers with zinc. Since the two stereoisomeric forms of a particular olefin usually have very close boiling points, it is difficult to separate them by distillation, and special methods have been developed for their synthesis. These include simultaneous dehydrohalogenation and de-

elastic polymers are based on homopolymers or copolymers of olefins and diolefins. Thus, polyethylene is made today by means of high-pressure peroxide-catalyzed processes or by low pressure methods using Ziegler-type catalysts Similar to polyethylene is polypropylene made by Ziegler-type catalysts. Polystyrene is made by emulsion

butyl rubber from isobutylene and butadiene or isoprene is made by means of aluminum chloride at low temperature. Polybutenes made by aluminum chloride or boron fluoride polymerization of isobutylene or isobutylene-butene mixtures vary in consistency from oils to rubbery polymers An exact duplicate of the stereochemical structure of *polyisoprene* in natural rubber has recently been made by using lithium catalysts. Synthetic drying oils have been made from butadiene and butadiene-styrene copolymers

Physical constants of unsaturated hydrocarbons

	Formula	Molecular weight	Boiling point, °C	Melting point, °C	Density, d_4^{20}	Refractive index, n_D^{20}
Monoolefins						
Ethylene	C_2H_4	28.0	−103 8	−169 4		
Propylene	C_3H_6	42.1	−47.7	−185.2	0.5139	
1-Butene	C_4H_8	56 1	−6.32	−185 4	0 5952	1.3777
cis-2-Butene	C_4H_8	56.1	3 64	−139 3	0.6213	1 3932
trans-2-Butene	C_4H_8	56.1	0 86	−105 8	0 6042	1 3842
Isobutene	C_4H_8	56 1	−6.93	−140 7	0 5942	1.3796
1-Pentene	C_5H_{10}	70.1	29 97	−138 0	0 6405	1.3715
cis-2-Pentene	C_5H_{10}	70.1	37.0	−179.0	0.6503	1.3822
trans-2-Pentene	C_5H_{10}	70 1	35 85	−135 5	0 6481	1.3785
2-Methyl-1-butene	C_5H_{10}	70 1	31.05		0 6504	1 3777
3-Methyl-1-butene	C_5H_{10}	70.1	18 8	−180 0	0 6332	1 3640
2-Methyl-2-butene	C_5H_{10}	70 0	38 49	−133.6	0.6620	1.3869
Cyclopentene	C_5H_8	68 1	44 1	−134 6	0 7716	1.4228
1-Hexene	C_6H_{12}	84 2	63 7	−141.0	0 6734	1.3880
Cyclohexene	C_6H_{10}	82 1	83 19	−103 6	0 8108	1 4467
Diolefins						
Propadiene	C_3H_4	40.1	−34 5	−136 1		
1,2-Butadiene	C_4H_6	54 1	10 3			
1,3-Butadiene	C_4H_6	54 1	−4 41	−108 9	0 6211	
1,2-Pentadiene	C_5H_8	68.1	44 9	−108 9	0 6206	1.4292
cis-1,3-Pentadiene	C_5H_8	68 1	44.2		0 6905	1.4359
trans-1,3-Pentadiene-1,3	C_5H_8	68 1	42 3		0 6764	1 4299
1,4-Pentadiene	C_5H_8	68 1	26 12	−148.8	0 6720	1 3880
3-Methyl-1,2-butadiene	C_5H_8	68.1	40.5	−120 0	0 6833	1 4166
2-Methyl-1,3-butadiene	C_5H_8	68 1	34 08	−146 8	0 6808	1 4216
Cyclopentadiene	C_5H_6	66 1	41.5	−85	0 8026	1.4429
1,3-Cyclohexadiene	C_6H_8	80 1	80 3	−104 8	0 8413	1.4740
2,3-Dimethylbutadiene	C_6H_{10}	82 1	68 5	−76 0	0 7263	1.4216
Acetylenes						
Acetylene	C_2H_2	26 0	−75	−85		
Methylacetylene	C_3H_4	40 1	−23 22	−102 7		
1-Butyne	C_4H_6	54 1	8 06	−125 7	0 6682	
2-Butyne	C_4H_6	54 1	26 99	−32 28	0 6910	1.3921
1-Pentyne	C_5H_8	68 1	40	−95	0 7221	1 4079
2-Pentyne	C_5H_8	68 1	56	−101	0 687	1 4004

using metallic sodium catalysts By far the largest single type of synthetic detergent marketed today (alkylaryl sulfonate) is derived from tetrapropylene made by the polymerization of propylene over a phosphoric acid catalyst.

Olefins and diolefins cyclize readily with a variety of reagents. Diolefins may form cyclic polymers or react with olefins to form cycloalkenes (*see* CYCLODODECATRIENE; CYCLOPENTADIENE; DIENE). Acetylenes cyclize either thermally to form aromatics or with the aid of catalysts to form polymeric cyclic compounds (*see* CYCLOOCTATETRAENE).

Addition reactions of olefins are among the most important in the entire field of organic chemistry. For a discussion of normal additions which follow Markownikoff's rule and abnormal additions made under the influence of oxygen or peroxides *see* ADDITION REACTION. Other addition reactions involve the alkylation of isoparaffins and aromatics with olefins. During World War II, much of the high-octane aviation gasoline was made by the alkylation of mixed isobutane-isopentane fractions which contained propylene and butenes. Sulfuric acid and hydrofluoric acid were used as catalysts A wide variety of alkylated aromatics are made by the alkylation of benzene with olefins using aluminum chloride, hydrogen fluoride at low or moderate temperature, or phosphoric acid on inert carriers at high temperature and pressure Thus, ethylbenzene is made on a tonnage basis and converted to styrene by catalytic dehydrogenation processes. Cumene for the production of phenol and acetone through the hydroperoxide is made by the catalytic alkylation of benzene with propylene Dodecylbenzene for conversion to sulfonated detergents is obtained mainly by the alkylation of benzene with tetrapropylene

Oxidation of olefins with air, oxygen, or peracids may occur at the double bond to give epoxides or with ozone to give ozonides which cleave under oxidizing or reducing conditions (*see* EPOXIDATION; OZONIZATION). With selected reagents such as selenium dioxide and lead tetraacetate, oxida

tion may take place at a methylene group adjacent to the double bond to give unsaturated ketones or unsaturated hydroxy esters

Sulfur, sulfur dioxide, sulfur trioxide, sulfuric acid, hydrogen sulfide, bisulfites, thiocyanogen, and sulfur monochloride add to olefins with varying degrees of ease to give a variety of useful products.

Halogens add to olefins across the double bond at low temperatures, but at high temperatures may give substitution products Thus, propylene when chlorinated at 500–600°C yields mainly allyl chloride Halogenated paraffins add to the double bond in the presence of Friedel Crafts catalysts. Perhalomethanes may add under the influence of free-radical reagents. Hypochlorous acid adds to olefins to yield chlorohydrins which may be converted to epoxides or glycols. Hydrogen adds normally to monoolefins under the influence of catalysts, but selective hydrogenation of diolefins may lead to 1,2 addition, 1,4 addition, or 3,4 addition Acetylenes may be selectively hydrogenated to the corresponding ethylenes by either catalytic or metal hydrogen donor hydrogenation. The "oxo" reaction for the synthesis of aldehydes and alcohols from olefins, carbon monoxide, and hydrogen is a commercially important reaction Over 50,000,000 lb of alcohols containing 4 to 13 carbon atoms per molecule, and made by the Oxo process in the United States is marketed annually [C.A.C.]

Uraninite

The chief ore mineral of *uranium* The element uranium was discovered *in the ore* by M H Klaproth in 1789 *Uraninite* has the idealized chemical composition UO_2, uranium dioxide Thorium and rare earths, chiefly cerium, are usually present in va[illegible] is [illegible] ni [illegible] te [illegible] (thorianite), and CeO_2 (cerianite) There are three chief natural sources of uraninite hydrothermal veins as in the Erz Gebirge of Saxony the Katanga district of the Belgian Congo, and Great Bear Lake Canada, essentially flat lying deposits in bedded sedimentary rocks chiefly conglomerates and sandstones, as in the plateau area of Colorado Utah and New Mexico, and pyritic conglomerate beds of Precambrian age, as in the Witwatersrand, Africa, and the Blind River area of Ontario *See* Radioactive minerals; Rare-earth elements; Thorianite, Uranium

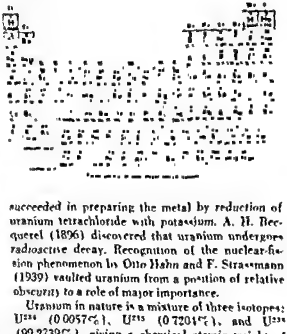
Uraninite is isometric in crystallization The crystal structure is of the calcium fluoride type Natural uraninite always contains U^{6+} in variable amounts, in addition to U^{4+}, and the formula may be written $(U_{1-x}^{4+}U_x^{6+})O_{2+x}$ The extra oxygen needed for valence compensation is housed interstitially in the vacant 8-coordinated positions of the crystal structure The mineral sometimes is found as cubic or octahedral crystals, especially in its occurrences in pegmatites It more often occurs as fine-grained masses or disseminations, or as crusts with a botryoidal surface and a layered, radial, fibrous to dense structure. The latter variety, also known under the name pitchblende, occurs chiefly in hydrothermal veins. The color of uraninite is black grading to brownish black and dark brown in the more highly oxidized material The luster of fresh material is steel gray. The hardness is 5½–6 on the Mohs scale. The specific gravity of pure UO_2 is 10.9, but that of most natural material is 9.7–7.5 *See* Lead isotopes, geochemistry of; Nuclear fuels; Ore and mineral deposits.

[C.FR.]

Bibliography: C. Frondel, *Systematic mineralogy of uranium and thorium*, USGS Bulletin 1064, 1958.

Uranium

Chemical element number 92, uranium, U, is a member of the actinide series, in which the 5*f* electron shell is being filled The element was isolated (1789) by M. H. Klaproth in a sample of pitchblende from Saxony. E. M. Peligot (1841) showed that the "semimetallic" element obtained by Klaproth was actually the dioxide. Peligot



succeeded in preparing the metal by reduction of uranium tetrachloride with potassium. A. H. Becquerel (1896) discovered that uranium undergoes radioactive decay. Recognition of the nuclear-fission phenomenon by Otto Hahn and F. Strassmann (1939) vaulted uranium from a position of relative obscurity to a role of major importance.

Uranium in nature is a mixture of three isotopes: U^{234} (0.0057%), U^{235} (0.7204%), and U^{238} (99.2739%), giving a chemical atomic weight of 238.07. U^{235} undergoes fission with slow neutrons to release large amounts of energy. U^{238} absorbs slow neutrons to form U^{239}, which in turn decays to fissile Pu^{239} by the emission of two β-particles. Other isotopes of uranium ranging in mass from U^{227} through U^{240} have been prepared by radioactive processes *See* Radioactivity.

Natural occurrence. Uranium is believed to be concentrated largely in the earth's crust, where the average concentration is 4 parts per million (ppm) For comparison, the earth's crust contains 0.1 ppm silver and 0.5 ppm mercury. Basic rocks (basalts) contain less than 1 ppm uranium, whereas acidic rocks (granites) may have 8 ppm or more.

Table 1. Uranium minerals

Mineral	Chemical composition	Color	Specific gravity	Typical occurrence
Uraninite	UO_2 (contains Th, rare earths)	Black	8–10.6	Arendal, Norway
Pitchblende (var.)	UO_{2+x}	Black	6–8	Belgian Congo
Euxenite-polycrase	$(Y,Ca,Ce,U,Th)(Nb,Ta,Ti)_2O_6$	Dark brown	4–6	Nipissing, Ontario
Samarskite	$(Y,Ca,Fe,U,Th)(Nb,Ta)_2O_6$	Black	5–6	Mitchell Co., N.C.
Davidite	$(Fe,Ce,U)(Ti,Fe,V,Cr)_3(O,OH)_7$	Black	4–5	Rum Jungle, Australia
Coffinite	$USiO_4$	Black	5–6	Colorado Plateau
Carnotite	$K_2(UO_2)_2(VO_4)_2 \cdot xH_2O$	Yellow	3–5	Colorado Plateau
Tyuyamunite	$Ca(UO_2)_2(VO_4)_2 \cdot xH_2O$	Yellow	3–4	Ferghana, Turkestan
Autunite	$Ca(UO_2)_2(PO_4)_2 \cdot xH_2O$	Greenish-yellow	3–4	Autun, France
Torbernite	$Cu(UO_2)_2(PO_4)_2 \cdot xH_2O$	Green	3–4	Erzgebirge, Saxony
Uranophane	$Ca(UO_2)_2Si_2O_7 \cdot xH_2O$	Greenish-yellow	3–4	Belgian Congo

Estimates for sedimentary rocks are 2 ppm, and for ocean water, 0.001 ppm. The total uranium content of the earth's crust to a depth of 15 miles is calculated to be 10^{14} tons, the oceans may contain 10^{10} tons of uranium.

Several hundred uranium-containing minerals have been identified, but only a few are of commercial importance. Table 1 summarizes data on some of the more important minerals. All uranium minerals contain lead, resulting from radioactive decay of uranium. Uraninite, as found in pegmatites, usually occurs in rather small amounts which are of little economic significance. Pitchblende, a variety of uraninite found in hydrothermal veins, is the most important mineral of uranium. It is usually poorly crystalline, contains very little thorium or rare earths, and is frequently found associated with sulfide minerals. Coffinite was first identified in 1951, but has since become recognized as an important uranium mineral on the Colorado Plateau. The euxenite-polycrase series, samarskite, and davidite are primary pegmatite minerals. The remaining compounds are secondary minerals formed by oxidation and weathering of primary uranium minerals. Near-surface uranium mineralization invariably consists of oxidized ore.

Prior to 1942, uranium was obtained principally as a by-product of radium mining operations. With the discovery of nuclear fission and the potential of atomic power, the possession of uranium reserves became of vital importance. Deposits containing as little as 0.1% uranium are being mined. Some of the largest occurrences are the hydrothermal vein deposits of Shinkolowbe, Belgian Congo, and at Great Bear Lake and Lake Athabaska in northwestern Canada. The vanadium-containing sedimentary deposits of the Colorado Plateau and the Blind River, Ontario, conglomerates are also important sources of uranium. In addition to these and numerous other occurrences, extensive reserves of low-grade ore (0.005–0.02% uranium) exist in phosphate deposits (Russia, Florida, and North Africa), bituminous shales (Russia, Sweden, and Tennessee), lignites (the Dakotas), and the gold reefs of the Witwatersrand (South Africa). *See* RADIOACTIVE MINERALS.

For a discussion of extraction methods, *see* URANIUM METALLURGY.

Uranium metal. Uranium is a very dense, strongly electropositive, reactive metal, ductile and malleable, but a poor conductor of electricity. It is most conveniently prepared by reduction of a halide (UF_4) with calcium or magnesium in a sealed vessel at 1300–1400°C. The steps involved in preparation of the metal from uranyl nitrate are summarized by the following equations:

$$UO_2(NO_3)_2 \cdot 6H_2O \xrightarrow{500°C} UO_3 + 2NO_2 + \tfrac{1}{2}O_2 + 6H_2O$$
$$UO_3 + H_2 \xrightarrow{700°C} UO_2 + H_2O$$
$$UO_2 + 4HF \xrightarrow{550°C} UF_4 + 2H_2O$$
$$UF_4 + 2Mg \xrightarrow{1300°C} U + 2MgF_2$$

Some of the physical and thermal properties of uranium are listed in Table 2. The metal exists in

Table 2. Physical and thermal properties of uranium

Property	Value
Melting point	1132 ± 1°C
Boiling point	3818°C
Vapor pressure, at 1630–1970°K	$\log p_{mm} = -\frac{2330}{T} + 8.583$
Heat of fusion	4.7 kcal/mole
Heat of vaporization	106.7 kcal/mole
Heat of sublimation, at 0°K	116.6 kcal/mole
Enthalpy, at 25°C	1521.4 cal/mole
Heat capacity, at 25°C	6.612 cal/(°C)(mole)
Entropy, at 25°C	11.99 cal/(°C)(mole)
Thermal conductivity, at 70°C	0.071 cal/(cm-sec)(°C)
Electrical conductivity	$2\text{–}4 \times 10^4$ ohm/cm

three crystalline modifications. α-Uranium (25–668°C) is orthorhombic ($a = 2.854$, $b = 5.869$, $c = 4.956$ A) with four atoms per unit cell and a density of 19.04. Its structure is interpreted as a distorted hexagonal lattice containing corrugated sheets of uranium atoms. The β phase (668–774°C) is a complex tetragonal structure ($a = 10.759$ A) with 30 atoms per cell and a density of 18.11 at 720°C. γ-Uranium (774–1132°C) is body-centered cubic ($a = 3.525$ A) with two atoms per cell. Its density is 18.06 at 805°C. The β form can be stabilized at room temperature by the addition of small amounts of chromium, the γ form with molybdenum.

The unique nature of the room-temperature, α structure curtails solid solution of uranium with many other metals. Extensive solid solution with-

Photomicrographs of aluminum-uranium alloy. Samples have been ground and polished by hand, then polished and anodized electrolytically. This produces a very thin oxide film, whose thickness is dependent on the substrata orientation. The grain size, location and amount of different phases, and other identifying characteristics of the microstructure are revealed by the color variations. (a) As cast. Dark discontinuous lines are UAl_4. (b) Cold-worked. The UAl_4 seen in (a) has been broken up rather severely, showing a highly deformed microstructure. (c) Recrystallized. Fine grains have nucleated after the cold-worked alloy has been heated above the recrystallization temperature. All ×75. (USAEC–Union Carbide Corp. Oak Ridge Natl. Lab.)

(a)

(b)

(c)

Table 1. Uranium minerals

Mineral	Chemical composition	Color	Specific gravity	Typical occurrence
Uraninite	UO_2 (contains Th, rare earths)	Black	8–10 6	Arendal, Norway
Pitchblende (var.)	UO_{2+x}	Black	6–8	Belgian Congo
Euxenite-polycrase	$(Y,Ca,Ce,U,Th)(Nb,Ta,Ti)_2O_6$	Dark brown	4–6	Nipissing, Ontario
Samarskite	$(Y,Ca,Fe,U,Th)(Nb,Ta)_2O_6$	Black	5–6	Mitchell Co., N C
Davidite	$(Fe,Ce,U)(Ti,Fe,V,Cr)_3(O,OH)_7$	Black	4–5	Rum Jungle, Australia
Coffinite	$USiO_4$	Black	5–6	Colorado Plateau
Carnotite	$K_2(UO_2)_2(VO_4)_2\ xH_2O$	Yellow	3–5	Colorado Plateau
Tyuyamunite	$Ca(UO_2)_2(VO_4)_2\ xH_2O$	Yellow	3–4	Ferghana, Turkestan
Autunite	$Ca(UO_2)_2(PO_4)_2\ xH_2O$	Greenish-yellow	3–4	Autun, France
Torbernite	$Cu(UO_2)_2(PO_4)_2\ xH_2O$	Green	3–4	Erzgebirge, Saxony
Uranophane	$Ca(UO_2)_2Si_2O_7\ xH_2O$	Greenish-yellow	3–4	Belgian Congo

Estimates for sedimentary rocks are 2 ppm, and for ocean water, 0 001 ppm The total uranium content of the earth's crust to a depth of 15 miles is calculated to be 10^{14} tons; the oceans may contain 10^{10} tons of uranium

Several hundred uranium-containing minerals have been identified, but only a few are of commercial importance. Table 1 summarizes data on some of the more important minerals. All uranium minerals contain lead, resulting from radioactive decay of uranium. Uraninite, as found in pegmatites, usually occurs in rather small amounts which are of little economic significance. Pitchblende, a variety of uraninite found in hydrothermal veins, is the most important mineral of uranium. It is usually poorly crystalline, contains very little thorium or rare earths, and is frequently found associated with sulfide minerals Coffinite was first identified in 1951, but has since become recognized as an important uranium mineral on the Colorado Plateau The euxenite-polycrase series, samarskite, and davidite are primary pegmatite minerals The remaining compounds are secondary minerals formed by oxidation and weathering of primary uranium minerals. Near-surface uranium mineralization invariably consists of oxidized ore.

Prior to 1942, uranium was obtained principally as a by-product of radium mining operations. With the discovery of nuclear fission and the potential of atomic power, the possession of uranium reserves became of vital importance. Deposits containing as little as 0.1% uranium are being mined Some of the largest occurrences are the hydrothermal vein deposits of Shinkolowbe, Belgian Congo, and at Great Bear Lake and Lake Athabaska in northwestern Canada. The vanadium-containing sedimentary deposits of the Colorado Plateau and the Blind River, Ontario, conglomerates are also important sources of uranium. In addition to these and numerous other occurrences, extensive reserves of low-grade ore (0 005–0.02% uranium) exist in phosphate deposits (Russia, Florida, and North Africa), bituminous shales (Russia, Sweden, and Tennessee), lignites (the Dakotas), and the gold reefs of the Witwatersrand (South Africa). *See* RADIOACTIVE MINERALS.

For a discussion of extraction methods, *see* URANIUM METALLURGY.

Uranium metal. Uranium is a very dense, strongly electropositive, reactive metal, ductile and malleable, but a poor conductor of electricity It is most conveniently prepared by reduction of a halide (UF_4) with calcium or magnesium in a sealed vessel at 1300–1400°C. The steps involved in preparation of the metal from uranyl nitrate are summarized by the following equations:

$$UO_2(NO_3)_2\cdot 6H_2O \xrightarrow{500°C} UO_3 + 2NO_2 + \tfrac{1}{2}O_2 + 6H_2O$$
$$UO_3 + H_2 \xrightarrow{700°C} UO_2 + H_2O$$
$$UO_2 + 4HF \xrightarrow{550°C} UF_4 + 2H_2O$$
$$UF_4 + 2Mg \xrightarrow{1300°C} U + 2MgF_2$$

Some of the physical and thermal properties of uranium are listed in Table 2. The metal exists in three crystalline modifications. α-Uranium (25–668°C) is orthorhombic ($a = 2.854$, $b = 5\,869$, $c = 4\,956$ A) with four atoms per unit cell and a density of 19 04 Its structure is interpreted as a distorted hexagonal lattice containing corrugated sheets of uranium atoms The β phase (668–774°C) is a complex tetragonal structure ($a = 10.759$ A) with 30 atoms per cell, and a density of 18.11 at 720°C γ-Uranium (774–1132°C) is body-centered cubic ($a = 3\,525$ A) with two atoms per cell Its density is 18 06 at 805°C The β form can be stabilized at room temperature by the addition of small amounts of chromium, the γ form with molybdenum

Table 2. Physical and thermal properties of uranium

Property	Value
Melting point	1132 ± 1°C
Boiling point	3818°C
Vapor pressure, at 1630–1970°K	$\log p_{mm} = -\frac{2330}{T} + 8\,583$
Heat of fusion	4 7 kcal/mole
Heat of vaporization	106 7 kcal/mole
Heat of sublimation, at 0°K	116 6 kcal/mole
Enthalpy, at 25°C	1521 4 cal/mole
Heat capacity, at 25°C	6 612 cal/(°C)(mole)
Entropy, at 25°C	11 99 cal/(°C)(mole)
Thermal conductivity, at 70°C	0 071 cal/(cm-sec)(°C)
Electrical conductivity	$2–4 \times 10^4$ ohm/cm

The unique nature of the room-temperature, α structure curtails solid solution of uranium with many other metals. Extensive solid solution with-

Photomicrographs of aluminum-uranium alloy. Samples have been ground and polished by hand, then polished and anodized electrolytically. This produces a very thin oxide film, whose thickness is dependent on the substrata orientation. The grain size, location and amount of different phases, and other identifying characteristics of the microstructure are revealed by the color variations. (a) As cast. Dark discontinuous lines are UAl. (b) Cold-worked. The UAl, seen in (a) has been broken up rather severely, showing a highly deformed microstructure. (c) Recrystallized. Fine grains have nucleated after the cold worked alloy has been heated above the recrystallization temperature. All X75. (USAEC–Union Carbide Corp. Oak Ridge Natl. Lab.)

(a)

(b)

(c)

years; the orbital velocity of 4 25 mi/sec; and the inclination of orbital plane to ecliptic of 0 8°.

The mean apparent equatorial diameter of its disk is about 3.7″, varying only slightly for a maximum of 3 9″ at opposition to a minimum of 3.5″ at conjunction. The corresponding linear equatorial diameter is about 30,000 miles, with an uncertainty of several per cent because of the smallness of the apparent disk. The polar flattening is of the same order as that of Jupiter or Saturn and can be directly seen under good conditions; the ellipticity is 0.083 or about 1/12 according to most direct measurements, but the motion of the apsides of the orbit of the first satellite suggests a smaller value, about 0.055 or 1/18. Thus the polar diameter may be about 7% less than the equatorial diameter, and consequently the volume is about 50 (Earth = 1) The mass, about 14.55 (Earth = 1), is well determined from the observed motions of the brighter satellites; the corresponding mean density is 1.56 g/cm^3, a low value characteristic of the four major planets. The mean surface gravity is about 9.4 m/sec^2, or very nearly the same as on Earth; however, the centrifugal force due to the rapid rotation reaches about 0 6 m/sec^2 at the equator, reducing the effective gravity to 8 8 m/sec^2.

The apparent visual magnitude at mean opposition, that is, when closest to Earth, is +5.8, so that Uranus is just visible to the naked eye when its exact position among the stars is known The corresponding value of the albedo is about 0 5, a high value characteristic of the cloud cover of the major planets, but somewhat uncertain due to possible errors in the adopted diameter.

The telescopic appearance of Uranus is that of a small, slightly elliptical blue-greenish disk strongly darkened near the limb. Occasionally faint bands or belts parallel to its equator can be seen when Earth is close to the equatorial plane, as was the case in 1882 and 1924 The direction of the polar axis is inclined only 8° to the planet's orbital plane, and the direction of rotation is retrograde with a period of about 10h 49m determined spectroscopically. When the polar axis is directed toward Earth, as in 1903 and 1945, no bands can be seen.

Spectroscopic observations indicate the presence in the atmosphere of methane (CH_4), the amount of gas present above the cloud layer being estimated at 1.5 km at standard temperature and pressure Any ammonia present must be in the solid state, because of the low temperature of 100°K prevailing at this distance from the Sun. As is the case with the other major planets, theoretical considerations indicate that hydrogen and helium must be the main gaseous constituents of the atmosphere This has been confirmed by the identification by G Herzberg in 1951 of an absorption band or line, which was discovered by G. P. Kuiper at 8270 A, as due to molecular hydrogen; the estimated partial pressure of hydrogen above the cloud layer is about 2 atmospheres.

Satellites. Uranus has five known satellites Four comparatively bright ones were discovered visually and one much fainter was found photographically. Their main elements are shown in the table. All the satellites are much too faint and too small to show a measurable disk; their linear diameters can only be estimated from their apparent brightnesses and an assumed value of the albedo which, by analogy with the inner satellites of Saturn, is taken to be about 0 8, that of water ice or ammoniacal snow With this value the diameters of the four main satellites of Uranus range from 400 to 1000 km

Satellites of Uranus

	Mean distance from Uranus, 10^3 miles	Sidereal period, days	Diameter, miles	Magnitude at mean opposition
I Ariel	119	2 520	380	(vis) 15 5
II Umbriel	166	4 144	250	(vis) 16
III Titania	273	8 706	620	(vis) 14 0
IV Oberon	364	13 463	500	(vis) 14 2
V Miranda	81	1 414		(phot) 17

The orbital planes of the satellites are close to the equatorial plane of the planet. Accordingly the orbits are sometimes seen edgewise, as in 1882 and 1924, when Earth was in this plane, and sometimes seen almost face on and appear nearly circular, when Earth is close to the polar axis of the planet, as was the case in 1903 and 1945. [C.D.S.]

Bibliography: H. N Russell, R S. Dugan, and J Q. Stewart. *Astronomy*, vol. 1, 1945

Telescopic appearance of Uranus

Urea

The diamide of carbonic acid Urea is a white, crystalline, water-soluble compound, with melting point 132 7°C and formula

$$H_2N-\overset{\overset{\displaystyle O}{\|}}{C}-NH_2$$

Important biologically as an end-product of normal animal and human protein metabolism, it is also

f synthetic value in urea-formaldehyde plastics, esides being used as a stabilizer for many explosives

Urea is made commercially from the partial ydrolysis of cyanamide,

$$H_2N{-}CN + H_2O \rightarrow H_2N{-}CO{-}NH_2$$

nd by heating carbon dioxide and ammonia nder pressure, whereby the following equilibrium s attained.

$$O_2 + NH_3 \rightleftharpoons HO{-}\overset{\overset{O}{\|}}{C}{-}NH_2 + NH_3 \rightleftharpoons$$

$$\overset{\oplus}{N}H_4\overset{\ominus}{O}{-}\overset{\overset{O}{\|}}{C}{-}NH_2 \rightleftharpoons H_2N{-}CO{-}NH_2 + H_2O$$

t can also be prepared from phosgene and excess mmonia,

$$Cl{-}\overset{\overset{O}{\|}}{C}{-}Cl + 4NH_3 \rightarrow 2NH_4Cl + H_2N{-}CO{-}NH_2$$

nd from the rearrangement of ammonium cyanate F Wohler, 1828).

Urea is used medically as a diuretic, and formerly in the treatment of indolent ulcers Because

...ith urea to form molecular complexes in which he molar ratio of hydrocarbon to urea is proportional to chain length. The complexes are easily lecomposed by water, thus, straight chain hydrocarbons may be separated from branched-chain somers

In the commercial production of urea-formaldehyde resins, two molecules of urea are condensed with three of formaldehyde using catalysts such as pyridine, ammonia, or hexamethylenetetramine. The condensation product, a syrup, is mixed with cellulose and coloring matter, and molded to the desired shape under pressure and elevated temperature. In textile treating, or in laminating operations, solutions of the resins are used

Both methylolurea, $H_2N{-}CO{-}NH{-}CH_2OH$, and dimethylolurea

$$HOCH_2{-}NH{-}CO{-}NH{-}CH_2OH$$

are known to be products of base-catalyzed condensations of urea and formaldehyde; it is likely that during the pressure-heating treatment, the hard, thermosetting resin is formed from progressive thre...

...UREA-FORMALDEHYDE-TYPE RESINS

[E B R.]

Urea (biochemistry)

The diamide of carbonic acid It is the main nitrogenous end product of the metabolic degradation of amino acids in mammals, amphibians, and chelonians (tortoises); it is also formed from ammonia, originating in the intestinal tract in the putrefaction of proteins The liver is the site of its formation, which involves a complicated sequence of reactions, with ornithine, citrulline, and arginine constituting the distinguishing members of a cyclic mechanism, Two molecules of ammonia and one molecule of carbon dioxide are converted to one molecule of urea with each turn of the cycle. The final reaction is catalyzed by arginase:

Arginine + Arginase = Urea + Ornithine

See AMINO ACIDS; LIVER; UREA. [H.H.MI.]

Urea-formaldehyde-type resins

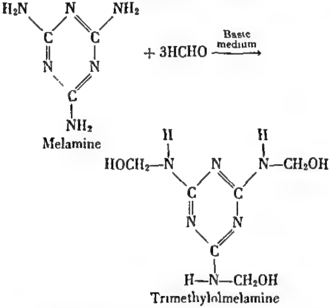
The condensation products obtained by the reaction of urea or melamine with formaldehyde Resinous condensation products of formaldehyde with other nitrogen-containing compounds, *for example*, aniline and amides, also belong to this group of resins but have gained only limited utility. Resins derived from the condensation of formaldehyde with urea, melamine, aniline, and *p*-toluene sulfonamide are discussed below

Urea- and melamine-formaldehyde resins. Because the amino resins possess an excellent combination of physical properties and can be easily fabricated in a variety of colors, they are widely used as wood glues, molding compounds, paper and textile finishes, and surface coatings. Methods of utilizing the resins are generally similar to the methods employed for several other condensation resins, such as the phenolic or epoxy resins First, intermediate condensation resins are prepared; then after ...

... by the reaction of urea or melamine with formaldehyde under neutral or mildly alkaline conditions to form mono-, di-, or polymethylol derivatives

$$\underset{\text{Urea}}{NH_2{-}\overset{\overset{O}{\|}}{C}{-}NH_2} + \underset{\text{Formaldehyde}}{HCHO} \xrightarrow[\text{medium}]{\text{Basic}} \underset{\text{Monomethylolurea}}{NH_2{-}\overset{\overset{O}{\|}}{C}{-}NH{-}CH_2OH}$$

$$\underset{\text{Urea}}{NH_2{-}\overset{\overset{O}{\|}}{C}{-}NH_2} + \underset{\text{Formaldehyde}}{2HCHO} \xrightarrow[\text{medium}]{\text{Basic}} \underset{\text{Dimethylolurea}}{HOCH_2{-}NH{-}\overset{\overset{O}{\|}}{C}{-}NH{-}CH_2OH}$$



Melamine

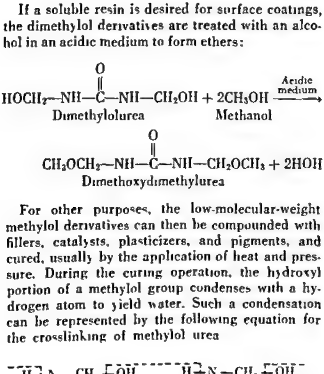
Trimethylolmelamine

If a soluble resin is desired for surface coatings, the dimethylol derivatives are treated with an alcohol in an acidic medium to form ethers:

$$\underset{\text{Dimethylolurea}}{HOCH_2{-}NH{-}\overset{\overset{\large O}{\|}}{C}{-}NH{-}CH_2OH} + \underset{\text{Methanol}}{2CH_3OH} \xrightarrow{\text{Acidic medium}}$$

$$\underset{\text{Dimethoxydimethylurea}}{CH_3OCH_2{-}NH{-}\overset{\overset{\large O}{\|}}{C}{-}NH{-}CH_2OCH_3} + 2HOH$$

For other purposes, the low-molecular-weight methylol derivatives can then be compounded with fillers, catalysts, plasticizers, and pigments, and cured, usually by the application of heat and pressure. During the curing operation, the hydroxyl portion of a methylol group condenses with a hydrogen atom to yield water. Such a condensation can be represented by the following equation for the crosslinking of methylol urea



Crosslinked polymer

Other reactions are also involved in curing. For example, the methylol groups can react with hydroxyl groups of alcohols (added as plasticizers), clays, or cellulose (added as fillers). Methylol groups can also react with each other to form ether linkages.

For immediate curing, a free acid can be used as a crosslinking catalyst. However, for the preparation of a stable molding powder that can be stored, a salt or ester that will liberate its corresponding acid at the high temperature used for molding usually serves as the catalyst.

As may be seen from the above equations, the completion of the curing reaction depends on the elimination of as much water as possible. Although most of the water is lost as vapor, a hygroscopic filler, such as cellulose, asbestos, or certain clays, is added to absorb the last traces of moisture. In addition, the filler can also react with some of the methylol groups of the resin intermediates.

Although the fabrication and applications of both the urea- and melamine-type resins are generally similar, there are several important differences in properties. The melamine type resins are more resistant to water, marring, and heat than the urea-type resins, but are, on the other hand, somewhat more expensive.

Curing in a mold results in translucent or opaque products, depending on the nature of the fillers and pigments used. The ease with which color can be introduced into amino resins has made it possible to make very attractive articles, such as dinner ware, buttons, appliance cases, handles, and knobs. In applications where resistance to heat and scratching is important, for example, in dinner ware, the melamine resins are preferred. *See* PLASTICS FABRICATION.

Another important application of both resins is in adhesives, especially for furniture and plywood. Depending on the resin and catalyst used, adhesives that can be cured under a variety of conditions may be formulated. Melamine resins are especially valuable as adhesives for laminates of paper or fabrics. *See* ADHESIVE; LUMBER MANUFACTURE.

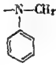
Soluble urea-formaldehyde and melamine-formaldehyde resins may be used to impregnate cloth or to treat paper. Curing of the resins in the cloth results in improved qualities, such as resistance to creasing or, in the case of melamine resins, in resistance to shrinkage in wools. Curing of the resins in paper results in an improvement in the strength of the paper when wet.

Although unmodified resins are common in the applications cited, modified resins or blends with other resins are often useful in some cases. In combination with alkyd resins, the ether derivatives, especially the ethers of the melamine resins, are the components in the formulation of baking enamels that are hard and resistant to water and detergents. Resins modified by the use of an alkyl-substituted urea or melamine are more flexible than the unmodified resin and are thus useful in coating compositions. Blends of urea or melamine-type resins with resins derived from resorcinol and formaldehyde are sometimes employed in adhesive compositions and as binders for the sawdust or wood chips used in the manufacture of particulate boards.

Another application of blends that is becoming steadily more important is the use of urea-type resins, together with polyvinyl alcohol, as impregnants for the fabric used in the manufacture of minimum-care or wash-and-wear cotton goods. See TEXTILE CHEMISTRY.

Aniline- and sulfonamide-formaldehyde resins. In addition to urea and melamine, other compounds containing $-NH_2$ groups can condense with formaldehyde to form methylol derivatives which are capable of further reaction. Although the corresponding resins have received limited attention, two examples are discussed here, aniline and p-toluenesulfonamide.

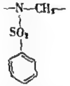
In a neutral medium, the reaction of aniline with formaldehyde yields a methylol derivative that exists as a cyclic trimer. When the trimer is heated, further condensation results in a resin with the following probable monomer unit:



Some crosslinking through the phenyl rings may also occur.

Because of their resistance to the absorption of water, the resins have been used in electrical applications, such as insulation or panels, where their natural brown color is not objectionable.

Similarly, resins can be prepared by curing methylol derivatives resulting from the condensation of p-toluenesulfonamide with formaldehyde. The probable monomer unit is

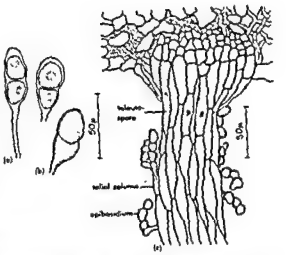


The resulting polymers are less colored than the aniline type resins, and have been employed in surface coatings. See CONDENSATION REACTION; FORMALDEHYDE; POLYMERIZATION; UREA.

[J.A.M., L.M.H.]

Uredinales

An order of parasitic fungi of the subclass Heterobasidiomycetidae which causes plant diseases known as rusts. It contains about 128 genera and 4600 species. The characteristic feature of the order is the teleutospore, a spore with one or more cells, each of which is a modified hypobasidium. Each cell may produce an epibasidium from which basidiospores are discharged. In the family Pucciniaceae, the teleutospores are stalked; in the Melampsoraceae they are laterally united to form crusts or columns, and in the third family, the Coleosporaceae, external epibasidia are lacking.



Teleutospore (resting spore). (a) Teleutospores of Uredinales. (b) Teleutospores of *Puccinia antirrhini* (snapdragon rust). (c) Telial column made up of teleutospores of *Cronartium ribicola* (white pine blister rust).

See HETEROBASIDIOMYCETIDAE; PLANT DISEASE; RUST (MICROBIOLOGY). [R.M.P.]

Bibliography. C. J. Alexopoulos, *Introductory Mycology*, 1952; G. M. Smith, *Cryptogamic Botany*, vol. 1, 2d ed., 1955.

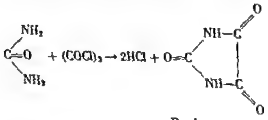
Ureid

An acyl derivative of urea, such as acetylurea, $CH_3CO-NHCONH_2$, and diacetylurea,

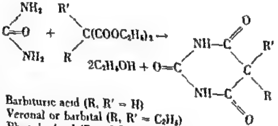
$$CH_3CO-NHCONH-COCH_3$$

The most important mem[illegible]

Parabanic acid is formed from urea and oxalyl chloride, barbituric acid or its derivatives (barbiturates) from urea and malonic ester or suitably substituted malonic esters (see ETHYL MALONATE).



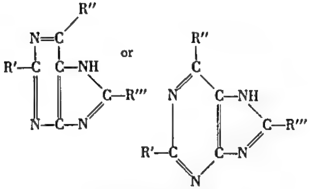
Parabanic acid



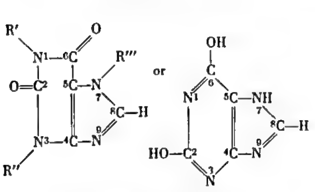
Barbituric acid (R, R′ = H)
Veronal or barbital (R, R′ = C_2H_5)
Phenobarbital (R = C_2H_5, R′ = C_6H_5)
Amytal (R = C_2H_5; R′ = isoamyl)

The barbiturates are acidic compounds, soluble in base (in which form they are often used). Introduced into medicine about 1900 (Veronal, 1903), the barbiturates are widely used in treating insomnia, epilepsy, hysteria, and as preliminary anesthetics. They should not be used indiscriminately because large doses are fatal, and continued use of small amounts often leads to habituation.

Many naturally occurring substances, for example, the purines (from nucleoproteins of plant and animal cells), and the xanthine alkaloids (from coffee, tea, and cocoa), contain diureid structures, in which two N—C—N units are combined with one C—C—C unit. Some examples follow:



Purine (R′, R″, R‴ = H)
Adenine, 6-aminopurine
(R′, R″ = H; R‴ = NH_2)
Guanine, 2-amino-6-hydroxypurine
(R′ = NH_2; R″ = OH, R‴ = H)
Uric acid, 2,6,8-trihydroxypurine
(R′, R″, R‴ = OH)
Xanthine, 2,6-dihydroxypurine
(R′, R″ = OH; R‴ = H)



Xanthine (R′, R″, R‴ = H)
Theophylline, 1,3-dimethylxanthine
(R′, R″ = CH_3; R‴ = H)
Theobromine, 3,7-dimethylxanthine
(R″, R‴ = CH_3; R′ = H)
Caffeine, 1,3,7-trimethylxanthine
(R′, R″, R‴ = CH_3)

See UREA. [E.B.R.]

Urethane

Urethane, or ethylurethane or ethyl carbamate, $H_2N—COOC_2H_5$, is the ethyl ester of carbamic acid. It forms white crystals with melting point 48–50°C, boiling point 182–184°C, and is freely soluble in water, alcohol, and chloroform. Less toxic than other urethanes, ethylurethane is a mild hypnotic and weak diuretic; in water, it forms a neutral solution with a cooling, saline taste. With quinine hydrochloride, it is used as a sclerosing agent for treatment of varicose veins.

Urethane is prepared commercially by two processes: urea and ethyl alcohol are heated under pressure,

$$H_2N—CO—NH_2 + C_2H_5OH \rightarrow NH_3 + H_2N—CO—OC_2H_5 \quad (1)$$

and urea nitrate is heated with ethyl alcohol and sodium nitrite,

$$H_2N—CO—NH_2{\cdot}HNO_3 + C_2H_5OH + NaNO_2 \rightarrow H_2N—CO—OC_2H_5 + NH_4NO_2 + NaNO_3 \quad (2)$$

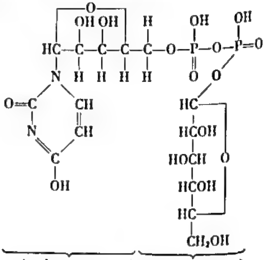
See AMIDE, ACID; RUBBER. [E.B.R.]

Uric acid

The chief nitrogenous end product in the metabolism of protein in insects, lizards, snakes, and birds, and the end product of purine metabolism in most insects, terrestrial reptiles, birds, man, and other primates. The purine bases are derivatives of a parent substance, purine, which contains the six-membered pyrimidine ring and the five-membered imidazole ring; they exist in all tissues in combination with phosphates in the form of nucleic acids, nucleotides, and polyphosphates. They are synthesized in the liver and kidneys from the products of protein metabolism. In mammals other than man and the primates, uric acid is oxidized to allantoin by the enzyme uricase, probably containing copper. *See* LIVER; NUCLEIC ACID; PROTEIN METABOLISM. [H.H.MI]

Uridine diphosphoglucose (UDPG)

A compound in which α-D-glucopyranose is esterified, at carbon atom 1, with the terminal phosphate group of uridine-5′-pyrophosphate. On very mild



(uridine portion) (glucosyl portion)
Uridine diphosphoglucose (UDPG)

acid hydrolysis, glucose and uridine-5′-pyrophosphate, or uridine diphospha x liberated Uri

dine diphosphoglucose occurs in animal, plant, and microbial cells and is synthesized enzymatically from uridine triphosphate (UTP) and α-glucose-1-phosphate by UDPG pyrophosphorylase. Inorganic pyrophosphate is liberated in the reaction:

$$\text{UTP} + \text{glucose-1-phosphate} \rightleftharpoons \text{UDPG} + \text{pyrophosphate}$$

Uridine diphosphoglucose is an intermediate in the metabolism of galactose. Galactose-1-phosphate reacts with UDPG to give glucose-1-phosphate and uridine diphosphogalactose. The last compound is isomerized to UDPG, which is thus regenerated in the process and then acts as a coenzyme in the conversion of phosphorylated galactose to glucose phosphates.

Uridine diphosphoglucose is also important in the synthesis of sucrose by plants:

$$\text{UDPG} + \text{fructose} \rightleftharpoons \text{uridine diphosphate} + \text{sucrose}$$

Through similar reactions, involving fructose-6-phosphate and glucose-6-phosphate in place of fructose, sucrose phosphate and trehalose phosphate are synthesized ...

See CARBOHYDRATE METABOLISM; GLUCOSE-1-PHOSPHATE [M.D.]

Urinalysis

The laboratory examination of urine. Examination of the urine requires a consideration of normal kidney and lower urinary tract anatomy and physiology, as well as the awareness of the clinical and laboratory conditions under which a urine sample is obtained and examined.

In general, urine is examined for color, specific gravity, pH, and the presence or absence of normal and abnormal materials. The color is often a rough index of the ability of the kidney to ... urine a... presence ... terial ... the spe... gravity varies between 1.015 and 1.030 under controlled conditions. Deviations above or below this range may indicate decrease in kidney function or the presence of some specific disease process. The pH reflects the role of the kidney in aiding in the maintenance of acid-base balance.

The presence of gross blood or microscopically discernible blood cells is almost always abnormal. The numbers and types of such cells may aid in diagnosis, especially when infection, trauma, or abnormal leaking of blood through the kidneys are encountered.

The sugar, or glucose, content of the urine bears a direct relationship to food intake normally and is often disturbed by disease processes, notably diabetes mellitus. Occasionally, rare types of sugars and related substances are present in persons suffering from certain familial or metabolic disorders. See DIABETES.

Since protein is usually conserved by the body, its presence in the urine may indicate either a general disease process in which there is an excessive protein excretion, or a specific kidney lesion in which the kidney is not able to retain protein. In addition, there are a few specific types of protein that are excreted in specific disorders. Examples include the Bence-Jones protein found in multiple myeloma, a form of blood tumor, and in several forms of albuminuria. See PROTEIN.

Casts are microscopic aggregates of different types which are formed in the renal tubules under some conditions of renal disease or damage. To the trained observer, these often afford excellent diagnostic clues concerning the nature of a patient's illness.

The bacteriologic examination of the urine is most important, particularly, of course, in cases where an infection is obvious or suspected. Visual examination, culture of the organisms found, and the use of sensitivity tests which permit the selection of effective antimicrobial agents are routine in such cases.

Other tests included in urinalysis are those of a chemical nature which have been devised to expose some material, normal or abnormal, that will aid in diagnosis, treatment, or prognosis.

Many of the above mentioned tests may be done in the physician's office, or at the bedside. Other more elaborate urinalyses may require special equipment, technicians, or at least the controlled regimen of a hospital examination which includes kidney function studies, radiography, and other more elaborate techniques. See CLINICAL PATHOLOGY. [E.G.ST.]

Urinary bladder

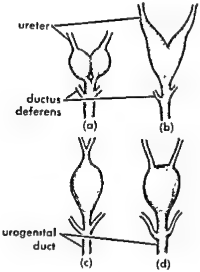
A distensible, muscular sac in most vertebrates which serves as a reservoir for urine. Snakes, crocodilians, birds (with the exception of the ostrich), most lizards, and a few fish lack a urinary bladder. In these organisms, urine empties directly into the cloaca. The development of the urinary system is intimately associated with the development of the reproductive system. Three general types of urinary bladder are recognized among the vertebrates: tubal, cloacal, and allantoic. See URINE.

Tubal bladders. Most fish possess tubal bladders, that is, enlargements of the mesonephric ducts. A single bladder, simplex type, results when two excretory ducts unite to form a single expanded structure. In the cod (*Gadus*) two separate bladders, duplex type, occur near the terminus of each excretory duct. These structures unite ...

Cloacal bladders. The cloacal bladder is found in monotremes, amphibians, and some dipnoans. There is no direct connection between the excretory ducts and this type of bladder. The bladder is an outpouching or diverticulum of the cloacal wall

The cloacal opening is closed by a sphincter muscle and the urine which seeps into the cloaca from the excretory ducts is forced into the bilobed bladder.

Allantoic bladder. This type of bladder is derived from the ventral wall of the cloaca and possibly the allantoic diverticulum. The role of the allantois in the formation of this type of bladder, which is found in most mammals, the turtles, and those lizards which have a bladder, is questioned by some embryologists. *See* ALLANTOIS.



Types of urinary bladders. (*a*) Duplex type. (*b*) Bilobed type. (*c*) Simplex type with ureters united. (*d*) Simplex type with separate ureters. (*From H. E. Walter and L. P. Sayles, Biology of the Vertebrates, 3d ed., Macmillan, 1949*)

The mammalian bladder is lined with a special epithelium composed of transitional cells. The muscular layer is composed of vertical, horizontal, and oblique fibers. The bladder drains through the urethra, the opening being controlled by a sphincter. Innervation is by the hypogastric sympathetic plexus and partly by parasympathetic fibers from the second and third sacral nerves. Stimulation of the parasympathetic causes the bladder muscle to contract and relaxes the internal sphincter. Micturition is a reflex act which is initiated voluntarily except in children. *See* PARASYMPATHETIC NERVOUS SYSTEM; SYMPATHETIC NERVOUS SYSTEM; URINARY SYSTEM; URINATION. [C.B.C.]

Urinary bladder disorders

The urinary bladder, vesica urinaria, is subject to anomalies, obstructions, inflammations, calculi, fistulae, and tumors. As the receptacle for urine from the kidneys it is variable in shape and form, depending upon the amount of urine it may contain.

Anomalies. A fairly common, distressing anomaly is exstrophy of the bladder, which is a failure of both the lower abdominal wall and anterior bladder wall to close. The urine appears directly through the abdominal wall defect. Occasionally a fetal structure, the urachus, fails to close by the time of birth and drains urine through the umbilicus. Abnormal insertions of the ureters are frequent but are important only if obstruction to urine flow is produced. Abnormal congenital valves of the posterior urethra may interfere with bladder emptying.

Obstruction. Interference with normal complete emptying of the bladder may be the result of various factors. In one form, an elevation of the internal urethral opening above its normal dependent position creates an unemptied pool of residual urine in the bottom of the bladder. This elevation is usually caused in the male by prostatic enlargement and in the female by cystocele, a bulging of the lower bladder wall into the vagina. Another form of obstruction is blockage of the bladder neck or urethra by such diverse lesions as anomalous congenital valves, prostatic enlargement, calculi, tumors or postgonorrheal urethral strictures. A third major cause of inadequate emptying is neurogenic, usually following spinal cord injury or disease. In this form the normal balance of bladder wall contraction and sphincter relaxation is lost, the bladder becomes greatly distended and then only partly empties itself.

Obstruction from any cause leads to hypertrophy of the muscle fibers. The bladder lining between these fibers may bulge outward and form pouches known as diverticuli. Other effects of obstruction include a susceptibility to infection and to stone formation, and back pressure on the ureters and kidneys. *See* KIDNEY DISORDERS.

Inflammation. Cystitis or inflammation of the urinary bladder is an extremely common affliction, annoying because of the accompanying painful frequent micturition. A major cause is the presence of residual urine after voiding, thus producing a good medium for bacterial growth. The organisms responsible are primarily colon bacilli and related gram negative rods, although enterococci and staphylococci are also frequent. Tubercle bacilli are rarely involved. Bacteria may gain access to the bladder from the blood after excretion by the kidneys, or may ascend from below through the urethra.

The lesions include granular or even cystic thickening of the epithelium, and redness and swelling of the wall.

Bladder calculi. Calculi usually develop as a result of infection and obstruction. They often reach a considerable size, up to several inches in diameter. They are usually composed of multiple urinary constituents, particularly phosphates and urates. Stones produce great pain by pressure in the sensitive trigonal area. After removal or spontaneous passage new stones tend to be formed unless the predisposing factors are removed.

Fistulae. Abnormal openings between the bladder or female urethra and the vagina are not unusual. Such fistulae most often follow injury during childbirth. The chief effects are infection and the constant leakage of urine through the vagina.

Tumors. The bladder is frequently the site of tumors which are almost always of transitional cell epithelial origin. The benign papilloma grows in the form of delicate fronds projecting into the bladder lumen. It is usually manifested first by hematuria, probably when a frond breaks off. It is often an indication of an abnormal bladder epithelium which continues to form new tumors. Malignant change into transitional-cell carcinoma is common. Such cancers may penetrate the bladder wall, invade neighboring structures, obstruct the ureters or urethra, or metastasize distantly.

Female urethra. This structure is not often diseased. In older women a painful raspberry-like swelling composed of inflammatory tissue may appear at the external urethral opening and is known as a caruncle. *See* REPRODUCTIVE SYSTEM DISORDERS. [R N B]

Urinary system

The urinary system consists of the kidneys, urinary ducts, and bladder.

COMPARATIVE EMBRYOLOGY

Kidney. The kidney, or nephros, of vertebrates is made up of many individual structural and functional units known as nephrons. The nephrons are derived from that portion of the embryonic mesoderm designated as the intermediate mesoderm. Ideally, this material becomes segmented, with each segment being termed a nephrotome (Fig. 1). Any given nephrotome contains a coelomic chamber, the nephrocoele, which opens to the adjacent body cavity via a peritoneal funnel.

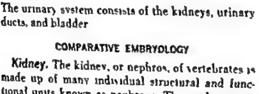


Fig. 1 Nephrotome of human embryo.

The conversion of nephrotome to nephron involves the following events. There arises from the dorso-later... growth, ... with the ... dial wal... nephrotome thins, flattens, and bulges inward coincidently with its invasion by arterial capillaries derived from the nearby dorsal aorta. The skein of capillaries comprises a glomerulus and the wall of the nephrotome investing the glomerulus is known as a renal, or Bowman's, capsule. The basic design of a nephron, as it results from these events, is shown in Fig. 2.

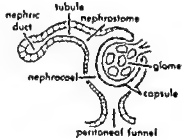


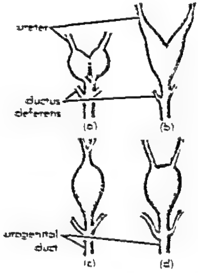
Fig. 2 Diagram of a nephron.

The original vertebrate nephros presumably consisted of similar nephrons throughout the length of the organ with each opening independently to the exterior. However, this arrangement is only hypothetical, for in all known vertebrates the nephrons empty into a common drainage, or nephric, duct which passes back alongside the nephros to terminate in the cloaca, the chamber which also receives the digestive tract. The situation is also complicated in present-day vertebrates by four variables. First, in the embryos of higher vertebrates, typical hollow nephrotomes are seldom formed; instead, nephrons differentiate without segmental arrangement within a continuous cord, the nephrogenic cord, of intermediate mesoderm. Second, as the nephros de... anterior ones tend to disappear before the posterior ones arise. Third, the nephrons become progressively more complex from anterior to posterior. Fourth, the manner of establishment of the nephric duct is not consistent in all vertebrates. The results of these four variables follow. *See* KIDNEY.

Nephros of fishes and amphibians. Embryonic development of the nephros is inaugurated within the most anterior reaches of the mesomere. This intermediate mesoderm becomes segmented into nephrotomes from each of which a nephron forms in the general manner already described. The number of these first formed nephrons varies with the species, but is always relatively small, usually 3–5. These nephrons, because of their anterior position and because they are the first to appear, comprise the head kidney, or pronephros. Accordingly, the nephrons themselves are termed pronephric tubules, or pronephrons. As the first and most anterior pronephric tubule arises, it first extends dorsolaterally and then turns backward to join the one forming immediately behind. This one in turn joins the one behind it, and so on, thus producing a common drainage duct designated the pronephric duct. The pronephric duct, once initiated, extends itself

The cloacal opening is closed by a sphincter muscle and the urine which seeps into the cloaca from the excretory ducts is forced into the bilobed bladder.

Allantoic bladder. This type of bladder is derived from the ventral wall of the cloaca and possibly the allantoic diverticulum. The role of the allantois in the formation of this type of bladder, which is found in most mammals, the turtles, and those lizards which have a bladder, is questioned by some embryologists. See ALLANTOIS.



Types of urinary bladders. (a) Duplex type. (b) Bilobed type. (c) Simplex type with ureters united. (d) Simplex type with separate ureters. (*From H. E. Walter and L. P. Sayles, Biology of the Vertebrates, 3d ed., Macmillan, 1949*)

The mammalian bladder is lined with a special epithelium composed of transitional cells. The muscular layer is composed of vertical, horizontal, and oblique fibers. The bladder drains through the urethra, the opening being controlled by a sphincter. Innervation is by the hypogastric sympathetic plexus and partly by parasympathetic fibers from the second and third sacral nerves. Stimulation of the parasympathetic causes the bladder muscle to contract and relaxes the internal sphincter. Micturition is a reflex act which is initiated voluntarily except in children. See PARASYMPATHETIC NERVOUS SYSTEM; SYMPATHETIC NERVOUS SYSTEM; URINARY SYSTEM; URINATION. [C.B.C.]

Urinary bladder disorders

The urinary bladder, vesica urinaria, is subject to anomalies, obstructions, inflammations, calculi, fistulae, and tumors. As the receptacle for urine from the kidneys it is variable in shape and form, depending upon the amount of urine it may contain.

Anomalies. A fairly common, distressing anomaly is exstrophy of the bladder, which is a failure of both the lower abdominal wall and anterior bladder wall to close. The urine appears directly through the abdominal wall defect. Occasionally a fetal structure, the urachus, fails to close by the time of birth and drains urine through the umbilicus. Abnormal insertions of the ureters are frequent but are important only if obstruction to urine flow is produced. Abnormal congenital valves of the posterior urethra may interfere with bladder emptying.

Obstruction. Interference with normal complete emptying of the bladder may be the result of various factors. In one form, an elevation of the internal urethral opening above its normal dependent position creates an unemptied pool of residual urine in the bottom of the bladder. This elevation is usually caused in the male by prostatic enlargement and in the female by cystocele, a bulging of the lower bladder wall into the vagina. Another form of obstruction is blockage of the bladder neck or urethra by such diverse lesions as anomalous congenital valves, prostatic enlargement, calculi, tumors or postgonorrheal urethral strictures. A third major cause of inadequate emptying is neurogenic, usually following spinal cord injury or disease. In this form the normal balance of bladder wall contraction and sphincter relaxation is lost, the bladder becomes greatly distended and then only partly empties itself.

Obstruction from any cause leads to hypertrophy of the muscle fibers. The bladder lining between these fibers may bulge outward and form pouches known as diverticuli. Other effects of obstruction include a susceptibility to infection and to stone formation, and back pressure on the ureters and kidneys. See KIDNEY DISORDERS.

Inflammation. Cystitis or inflammation of the urinary bladder is an extremely common affliction, annoying because of the accompanying painful frequent micturition. A major cause is the presence of residual urine after voiding, thus producing a good medium for bacterial growth. The organisms responsible are primarily colon bacilli and related gram negative rods, although enterococci and staphylococci are also frequent. Tubercle bacilli are rarely involved. Bacteria may gain access to the bladder from the blood after excretion by the kidneys, or may ascend from below through the urethra.

The lesions include granular or even cystic thickening of the epithelium, and redness and swelling of the wall.

Bladder calculi. Calculi usually develop as a result of infection and obstruction. They often reach a considerable size, up to several inches in diameter. They are usually composed of multiple urinary constituents, particularly phosphates and urates. Stones produce great pain by pressure in the sensitive trigonal area. After removal or spontaneous passage new stones tend to be formed unless the predisposing factors are removed.

Fistulae. Abnormal openings between the bladder or female urethra and the vagina are not unusual. Such fistulae most often follow injury during childbirth. The chief effects are infection and the constant leakage of urine through the vagina.

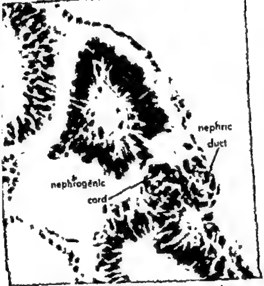
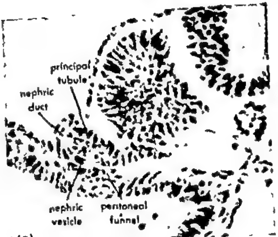


Fig 5 Origin of nephric duct in human embryo.



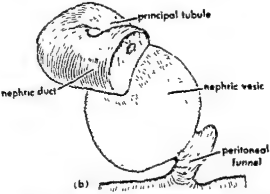


Fig 6 Early human mesonephron. (a) Section (b) Reconstruction

one category of animals to another, but in general they conform to the pattern described above. The human embryo may be used as a specific illustration Briefly, the nephrogenic cord paralleling the growing Wolffian duct provides an increasing number of serially arranged spherical bodies known as nephric vesicles (Fig 6) These are solid at first, but become hollow, and as they do so each vesicle sends a principal tubule dorsolaterally to join the Wolffian duct The vesicle proper will provide the capsule surrounding the later developing glomerulus Elongation and twisting of the principal tubule and the acquisition of a capillary network completes the nephron

Wolffian body These nephrons are known as mesonephric tubules, or mesonephrons, and collectively they comprise the mesonephros, or Wolffian body This kidney is well developed in reptiles and birds, but in mammals exhibits considerable variation In the rat embryo, for example, it is quite rudimentary and only a dozen or so abortive nephrons arise. At the other extreme is the pig embryo whose mesonephros is large and bulky and involves several hundred long and convoluted nephrons The human mesonephros lies between these two extremes

Embryonic function The variable status of the mesonephros, especially in mammals, raises the interesting question of the functional role it plays in the economy of the embryo It is reasonable, of course, to infer function from exhibited structure More convincing, however, are the direct experimental demonstrations of functional capacities of the mesonephrons. To this end, certain of the techniques employed to assess kidney functions in adult forms have been applied with profit.

The chick embryo has bee...
for ...
blue ...
appear in the mesonephrons. Another approach has been that of cultivating fragments of mesonephros in a suitable culture medium to which an indicator such as phenol red has been added The proximal portions of the tubules pick up the indicator and transport it to their lumina; the distal portions resorb water Still another method has been the direct identification of nitrogenous wastes deposited in the embryonic bladder, the allantois *See* ALLANTOIS.

Similar experiments on mammals are complicated by the intrauterine location of the embryo and its association with a placenta Nevertheless, it

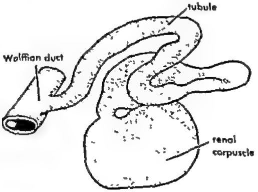


Fig 7 Diagram of later human *mesonephron*

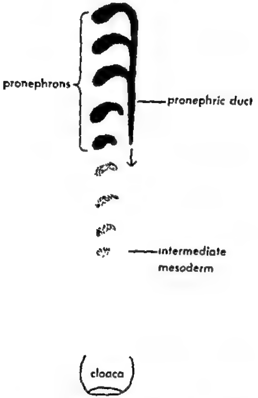


Fig. 3. Diagram of developing pronephros

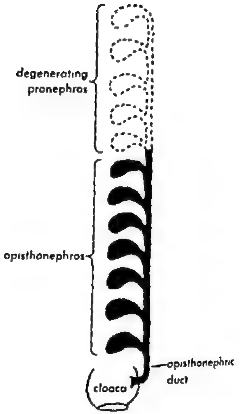


Fig. 4. Diagram of opisthonephros.

backward along the still undifferentiated intermediate mesoderm until it joins the cloaca (Fig. 3).

With the exception of the hagfishes and some bony fishes in which it persists throughout life, the pronephros has only a temporary existence. In the sharks, for instance, it is present only during early embryonic stages and has little or no functional significance, but in those forms having an active free-living larval stage, such as amphibians and lampreys, it persists in the larva as a functional organ and the individual pronephrons possess ciliated peritoneal funnels opening to the adjacent coelom. This provides for the uptake of certain materials directly from the coelom as well as from the blood stream Direct demonstrations of the functional capacities of pronephric tubules have come from a variety of tests. For example, the pronephrons of the larval lamprey will take up quantities of colloidal carbon which may be injected into the coelom; and through their ability to accumulate the dye phenol red, the pronephrons of frog larvae and certain bony fishes have also revealed their functional capabilities.

Whatever the length of its existence, the pronephros is supplemented and succeeded by a second generation of nephrons derived from the remainder of the intermediate mesoderm. Although they arise in basically the same manner and exhibit the same fundamental structure, these later-formed nephrons tend to be longer, more complex in their make-up, and ordinarily lack peritoneal funnels. Unlike their pronephric forerunners these nephrons fail to establish their own drainage duct; they join the already existing pronephric duct. Eventually, as noted, the earlier-formed pronephros disappears, leaving this later generation of nephrons to constitute the final, definitive kidney. This organ, distinctive to fishes and amphibians, is known as the opisthonephros, or back kidney, and the onetime pronephric duct which its nephrons have taken over is termed the opisthonephric duct (Fig 4).

Nephros of reptiles, birds, and mammals. The embryonic initiation of the nephros of higher vertebrates is customarily described as involving the establishment of a pronephros and pronephric duct as in fishes and amphibians. Although this may be true for reptiles and birds, in mammals pronephric tubules rarely if ever appear. In man, for example, that level of intermediate mesoderm equivalent to the pronephros never gets beyond the point of conversion to a few rudimentary nephrotomes. Because definitive pronephric tubules are never provided, the pronephric duct must arise in some fashion other than by junction of the ends of pronephrons Instead, the original nephric duct is initiated as a solid rod which splits off the dorsolateral side of the nephrogenic cord (Fig. 5). Once established in this fashion, the solid duct then frees itself from the parent material and as a tapering rod extends itself backward by independent terminal growth, ultimately contacting and fusing with the wall of the cloaca The solid rod gradually hollows out and within a few days after its original establishment becomes tubular throughout.

Wolffian duct. As the nephric duct, or Wolffian duct, is being formed the first of two generations of nephrons appears The first generation consists of a series of nephrons derived from the middle level of the intermediate mesoderm. The details of their manner of development and [illegible] vary from

gresses and the metanephros assumes full responsibility for excretion In females, the mesonephros disappears almost entirely; in males, parts of it are incorporated in the reproductive system. As a consequence of straightening and elongation of the fetal body, the kidneys come to be displaced relatively far forward in the body. *See* REPRODUCTIVE SYSTEM

Urinary bladder. At or near the posterior ends of the nephric ducts there frequently is a reservoir for urine This is the urinary bladder Actually there are two basic varieties of bladders in vertebrates. One is found in fishes in which the reservoir is no more than an enlargement of the posterior end of each urinary duct Frequently the urinary ducts are conjoined and a small bladder is formed by expansion of the common duct The far more common type of bladder is that exhibited by tetrapods. This is a sac which originates embryonically as an outgrowth from the ventral side of the cloaca Present in all during embryonic life, it is exhibited differentially in adults All amphibians retain the bladder, but it is lacking in snakes, crocodilians, and a few lizards, birds, also, with one or two exceptions, lack a bladder It is present in all mammals Because much of the developmental history of the tetrapod bladder is linked to the history of the cloaca, it will be considered in this conjunction

Amphibians The basic pattern is exemplified by the Amphibia All retain the cloaca as adults and the urinary bladder arises as a diverticulum from the floor of the cloaca Commonly it is bilobed There is no direct connection between the excretory ducts and the bladder Instead, urine first passes into the cloaca and thence into the bladder Urine is expelled by the intermittent opening of the cloacal orifice and the coincidental contractions of the muscular wall of the bladder

Reptiles and birds In reptiles and birds the cloaca is partly subdivided so that the intestine and urogenital ducts open into separate compartments which then join in a common outlet In the embryo a pouch develops from the floor of the urogenital portion of the cloaca and expands to form a prominent sac known as the allantois *See* EXTRAEMBRYONIC MEMBRANES. Ultimately the allantois enlarges to extend beyond the confines of

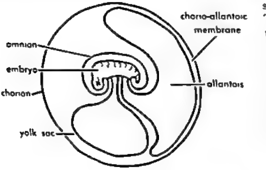


Fig. 10. Diagram of extraembryonic membranes of

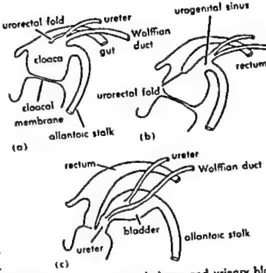


Fig 11. Development of cloaca and urinary bladder. (a) Formation of cloacal membrane (b) Urorectal fold formation (c) Urethra formation

the embryo and to fuse broadly with the outermost membrane, the chorion, surrounding the embryo (Fig 10) The combined chorioallantoic membrane ultimately makes broad contact with the membrane lining the inner surface of the shell. This fusion and spreading of the allantois serves to bring an extensive blood supply adjacent to the shell and thus sets up a medium for respiratory exchange. The allantois, therefore, serves not only as a reservoir for urine but also plays a major role in embryonic respiration.

[illegible] allantois is discarded at hatch- [illegible]

the turtles and certain lizards, retain the base of the allantois as an adult bladder and the remainder atrophies

Mammals Among mammals, only a few primitive forms retain a cloaca as adults In all the others it is modified in such a way as to be eliminated Concomitantly, the openings of the excretory ducts are shifted and the urinary bladder is established. The pig embryo may serve to illustrate the usual events.

Almost as soon as the cloaca is established, an allantois arises from its floor. As in reptiles and birds, the allantoic sac, complete with blood vessels, expands greatly and fuses with the chorion This association enables it to serve not only as a urinary bladder, but also to provide a major component of the placenta, so important in fetal economy In the meantime the cloaca itself is modified. The cloaca initially ends blindly, its ventro-posterior floor making contact with the embryonic skin to form the cloacal membrane (Fig 11*a*) A division of the cloaca into two parts, a dorsal rectum and ventral urogenital sinus, then follows This division is effected by the urorectal fold, a crescentic fold which works backward from the angle where the allantois, excretory ducts, and gut meet until it meets the cloacal membrane (Fig. 11*b*, *c*) Not only is the cloaca itself subdivided, but the

has been possible either to inject suitable indicators directly into the embryonic body or to introduce them secondarily by transmission via the placenta from the maternal blood stream. Such tests on embryos of the rabbit, cat, pig, and pouch-young opossum have provided positive demonstrations of the functioning of both the glomerular filter and the tubule proper.

Metanephros. As the mesonephros is attaining its maximum development, a second generation of nephrons is inaugurated. The history of this group is considerably more complicated than the former and runs briefly as follows. A tubular outgrowth from the Wolffian duct, the ureteric diverticulum, appears close to its entrance to the cloaca and pushes itself anteriorly into the still undifferentiated mesomere behind the mesonephros. The distal end of this diverticulum enlarges and the mesomere coincidently begins to condense around the enlargement Figure 8*a* shows that the proximal segment of the original diverticulum is the ureter, or metanephric duct; the expanded distal end of the diverticulum is the primitive renal pelvis; the condensed mesomere around the pelvis is the metanephric blastema

Subsequent events pertain primarily to the pelvis and blastema. The former first subdivides to form the two future major calyces (Fig. 8*b*). These divide and subdivide until several generations of branches are produced. The earlier generations come to represent the minor calyces; the later generations become the collecting tubules (Fig 8*c*, *d*). As the primitive pelvis carries on this program of subdivision, the blastemal tissue subdivides into a corresponding number of masses.

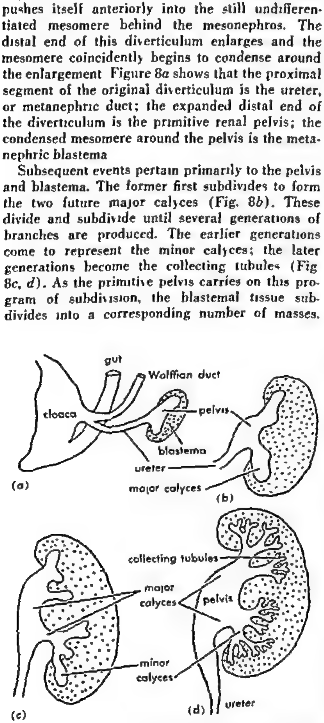


Fig 8. Stages in development of metanephric pelvis and blastema. (a) Ureteric diverticulum. (b) Major calyces. (c) Minor calyces. (d) Collecting tubules.

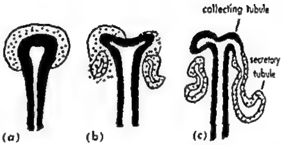


Fig. 9. Stages in origin of metanephric tubule (a) Vesicle formation. (b) Tubule formation. (c) Union of tubule and duct

Thus, there results a nodule of blastema in association with the end of each prospective collecting tubule (Fig. 9*c*). Each such nodule or sphere is the forerunner of a metanephric or uriniferous tubule. The solid sphere first becomes a vesicle which, by elongation, is transformed into a tortuous tubule (Fig. 9*a*, *b*, *c*). The thinner-walled, blind end of the tubule becomes the capsule surrounding the concomitantly forming glomerulus Coincidently, the tip of the prospective collecting duct grows out to meet the end of the tubule and the two unite (Fig 9*c*). The uriniferous tubules, collecting tubules, and calyces comprise the definitive kidney or metanephros of the late embryo and adult. In terms of gross anatomy, the convoluted portions of the uriniferous tubules, with their glomerular capsules, collectively comprise the cortex of the kidney These tubules lead into the collecting system and calyces making up the medulla of the kidney. All drainage ultimately converges upon the renal pelvis which in turn leads to the ureter.

Developmental interdependence. Experimental analyses have revealed important developmental interdependencies within the mesonephric and metanephric systems. The original nephric duct grows back from the level of its origin to join the cloaca. In the normal course of events it is joined along the way by mesonephrons and thus becomes the Wolffian duct If the backward extension of the duct alongside the prospective mesonephros-forming area is prevented, little or no development of the mesonephros will occur This can be accomplished readily in the chick embryo by producing a minute wound or inserting some block in the pathway of the duct. This is interpreted to mean that the duct serves as an inductor of the mesonephric tubules, that is, differentiation of the mesonephrons depends upon some kind of stimulus provided by the nephric duct. Another consequence of such a blockage of the nephric duct is the elimination of the ureter which normally grows from it This in turn leads to a failure of the blastema to produce metanephric tubules The ureter is the inductor of these tubules. Development of the nephros is thus revealed as a series of steps, each dependent upon the one before, with the original nephric duct playing the starting role.

During the latter part of embryonic life the mesonephros and metanephros n simultaneously. Gradually, however, hros

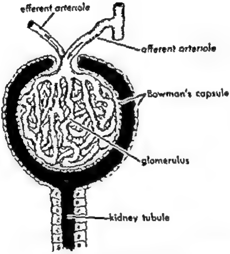


Fig 13 A renal corpuscle (From C. K. Weichert, *Anatomy of the Chordates*, 2d ed., McGraw-Hill, 1958)

glomeruli are typically present. These are small knots of arterial vessels, each surrounded by a double-walled cup called Bowman's capsule, the two together being known as a renal corpuscle (Fig 13) In some forms the anterior tubules of the opisthonephros lie in the same segments as posterior pronephric tubules. This indicates the transitional nature of the two A typical opisthonephric tubule consists of a narrow neck adjacent to the renal corpuscle, followed in turn by secretory and collecting portions. The collecting portion joins the archinephric duct. The ends of several collecting tubules may unite to form a common duct which either opens into the archinephric duct or establishes an independent connection with the cloaca Several such accessory ducts may be present.

Cyclostomes The opisthonephros of the adult hagfish differs basically from the original archinephros only in the loss of peritoneal connections of the posterior tubules. In the adult lamprey the opisthonephros on each side consists of a long strap-shaped ... The ... line ... like ... the ... of the kidney The ducts of the two sides unite posteriorly to open into a urogenital sinus which leads to the outside through an aperture at the tip of a small urogenital papilla Two slitlike genital pores connect the urogenital sinus with the coelom The condition is similar in both sexes Eggs or spermatozoa leave the body cavity via the genital pores, urogenital sinus, and urogenital aperture Only here are the reproduction and urinary systems associated

Fishes There is much variation in shape of the opisthonephric kidneys of fishes but they are fundamentally similar in structure. In some they extend the length of the coelom, in others they are short and may show various degrees of fusion Peritoneal funnels rarely occur Some marine teleosts lack glomeruli and thus possess aglomerular kidneys. In elasmobranchs the anterior ends of the kidneys of the male have been appropriated by the reproductive system Modified kidney tubules, called efferent ductules, connect each testis with the archinephric duct which lies on the ventral surface of the kidney and serves as a ductus deferens for sperm transport. Accessory urinary ducts are usually present. In teleost fishes there is no connection between the testes and the opisthonephric kidneys. The posterior ends of the archinephric ducts of female fishes enter a common urinary sinus inside a small urinary papilla. The latter enters the cloaca in elasmobranchs and dipnoans, but in most other fishes it opens directly to the outside, a cloaca being absent.

Amphibians. In common with other amphibians, adult caecilians possess an opisthonephros. The kidneys of the tailed amphibians are much like those of elasmobranch fishes, the anterior ends in males being concerned with genital rather than urinary functions (Fig. 14). The archinephric duct courses along the lateral edge of the kidney a short distance outside the kidney proper. Numerous collecting ducts or tubules leave the kidney at intervals to join the duct. The two ducts in both sexes open separately into the cloaca In frogs and toads the opisthonephric kidneys lie toward the posterior part of the abdominal cavity. A yellowish adrenal gland is located along the ventral surface of each. The kidneys of females are not related to the reproduction system but in males an intimate connection exists. Modified tubules, or efferent ductules, from the testes connect with kidney tubules which lead to the archinephric duct. This duct is located

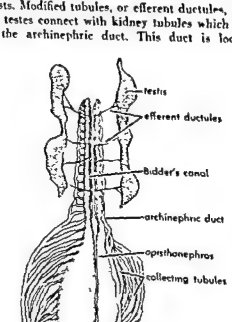


Fig 14 Urogenital organs of male salamander, ventral view. The collecting ducts on the right side are shown detached from the cloaca and spread out for the sake of clarity. (From C. K. Weichert, *Anatomy of the Chordates*, 2d ed., McGraw-Hill, ...

cloacal membrane is likewise divided into an anal and urethral membrane. Both these membranes eventually rupture so that the two parts of the cloaca open independently to the exterior.

The next step is the enlargement of that portion of the urogenital sinus receiving the neck of the allantois. The enlargement may involve a part of the allantoic stalk itself. This enlargement is the beginning of the definitive urinary bladder. As the bladder expands, the terminal ends of the mesonephric ducts which open to the urogenital sinus at this level are absorbed into the bladder wall. In consequence, the ureters, which stemmed initially from the Wolffian ducts, come to open directly to the growing bladder while the Wolffian ducts open somewhat behind into that part of the sinus which remains narrower and gives rise to the urethra (Fig. 11c). The ultimate fate of the Wolffian ducts and the final form of the urethra differ between the two sexes. The allantois per se is discarded at birth. [T.W.T.]

COMPARATIVE ANATOMY

Similarities are not particularly evident among the many and varied types of excretory organs found among vertebrates The variations that are encountered are undoubtedly related to problems with which vertebrates have had to cope in the past in adapting themselves to different environmental conditions under which they have lived

Archinephros. It is now generally believed that the primitive vertebrate ancestor possessed an excretory organ referred to as an archinephros or holonephros (Fig 12). This probably consisted of a pair of dorsally located ducts extending the length of the body cavity. Each duct was joined by a series of segmentally arranged tubules, one pair to each segment The other end of each tubule opened into the body cavity by a ciliated, funnel-shaped aperture. Close to each opening was a small knot of arterial blood vessels called an external glomerulus. From this type of kidney with its archinephric duct the various kidneys of forms living today may originally have been derived The larval form of the hagfish and the larvae of certain amphibians, the caecilians, are present-day vertebrates possessing kidneys of this type.

Anamniote kidneys. The anterior portion of the archinephric kidney persists only in the adult stage of the hagfish and of certain teleost fishes in which it is called the head kidney, or pronephros Nevertheless, it appears in the embryos of most vertebrates as a transitory structure. This usually degenerates soon after it has formed. *The remainder* of the kidney posterior to the pronephros is known as the opisthonephros.

Pronephros. The importance of the pronephros lies mainly in the part it plays during development in forming the archinephric duct which persists even though the pronephros appears only as a transient structure. In some larval forms the pronephros may be important in getting rid of wastes at a time when the opisthonephros is being formed. Even in the hagfish, the head kidney has

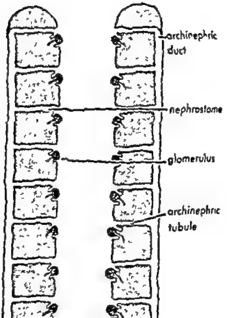


Fig. 12 Diagram showing hypothetical structure o[f] archinephros (From C K Weichert, Anatomy of th[e] Chordates, 2d ed., McGraw-Hill, 1958)

become modified from the primitive condition Her[e] the openings of the tubules connect with the peri[-] cardial cavity and the fluid drained by the tubule[s] passes into a nearby vein instead of entering th[e] archinephric duct Pronephros and opisthonephro[s] become completely separated by degeneration of the portion between them

Opisthonephros. Because the pronephros is usu[-]ally a transitory structure, the opisthonephros is the more important of the two. It serves as the adult kidney in lampreys, most fishes, and amphib[-]ians The opisthonephros differs from the proneph[-]ros in several respects. A main distinction is that the segmental arrangement of the kidney tubules is lost and many tubules may lie within the confines of a single segment. Furthermore the connection of the opisthonephric tubules with the body cavity is usually lost and renal cor[illegible] intern

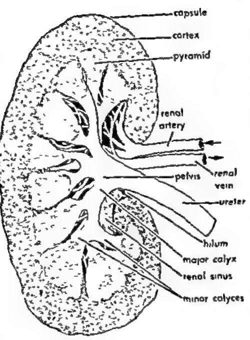


Fig 16 Sagittal section of metanephric kidney of man (semidiagrammatic). (From C K Weichert, *Anatomy of the Chordates*, 2d ed , McGraw-Hill, 1958)

capsule of connective tissue under which lies the cortex The renal corpuscles and the greater part of the secretory tubules lie entirely in the cortex The portion of the kidney surrounded by the cortex is the medulla It is partly composed of large areas, the renal pyramids The outer borders of the pyramids are divided into smaller units called lobules The collecting tubules lie within the pyramids but may extend well up into the cortex. The inner portion of each pyramid, in the form of a blunt papilla, projects into an outpocketing of the pelvis known as a minor calyx Several minor calyces join together to enter major calyces which in turn open into the renal pelvis The pelvis leads to the ureter which empties into the bladder except in monotremes in which it enters the urethra Urine, which is stored temporarily in the bladder passes to the outside through the urethra In males the urethra opens at the tip of the penis, in females the condition varies for in some, as in the rat and mouse the urethra opens independently to the outside passing through the clitoris It usually, however, enters a vestibule which is the terminal part of the genital tract The kidneys of mammals are markedly lobulated in the embryo and in many forms this condition is retained throughout life. [C K W]

COMPARATIVE VERTEBRATE PHYSIOLOGY

The function of the urinary system is to excrete metabolic waste products, primarily water, salts, and nitrogenous compounds, from the organism in a regulated manner so that the composition of the body fluids remains nearly constant in spite of variable intake The capacity and efficiency of the system differ from animal to animal and are related to the demands placed on the system. For example, a fresh-water animal excretes large amounts of water with little salt, whereas a desert animal must conserve water and excrete salts in high concentrations. The observed differences in *function* are valuable in furthering the general understanding of how the urinary system operates in all animals, including man.

Urine production. This is a continuous process that takes place in the kidneys. The rate of flow depends on water intake and diet and can vary over a wide range in the individual animal, about 50-fold in mammals. Different animal groups have different average rates of urine flow according to their type of habitat (Fig. 17)

Physical properties. In fish and amphibians, the urine is a clear, colorless, watery fluid. In birds and many reptiles, the urine, because of a high content of uric acid is a semisolid white paste In mammals, it is usually clear and yellow. The freezing point depression, Δ°C, and the specific gravity of [illegible] of the concentration of sol[illegible]

urine is always more dilute than the blood The Δ°C may vary from values as low as 0 05°C to almost the value of the blood (Fig 17). In birds, the Δ°C of the urine may be slightly higher than the blood In mammals, it varies with the water intake over a wide range In certain mammals, who must usually conserve water, it may reach values as high as 20 times the blood, the Δ°C = 10.

Chemical composition The nitrogenous compounds excreted in the urine are mainly ammonia, trimethylamine oxide (TMAO), urea, uric acid, creatine and creatinine The principal salts are chlorides, phosphates, sulfates, and bicarbonates of sodium and potassium In addition to the normal waste products of metabolism, many substances foreign to the body will be excreted.

Most nonmammalian vertebrates living in fresh water such as fresh-water fishes, aquatic amphibians and alligators, excrete nitrogen in the easily diffusible but poisonous form of ammonia In marine fish nitrogen is excreted mostly as urea, TMAO and creatine Most amphibians shift at metamorphosis from ammonia to urea excretion Many reptiles excrete both urea and uric acid Birds excrete predominantly uric acid and mammals almost exclusively urea

Renal function theories. It was suggested by K. Ludwig (1844) and later by A Cushny (1917) that urine formation starts in the glomeruli where a protein-free filtrate of the blood is formed. The theory was experimentally confirmed by A. Richards A Walker and collaborators (1927–1938) who showed that the glomerular fluid has the same chemical composition as a true ultrafiltrate of the blood

In the renal tubules the filtrate undergoes significant changes Water, salt, and other substances are reabsorbed from the tubule back into the blood

within the kidney along its lateral margin. Peritoneal funnels are present on the ventral sides of the kidneys in some frogs but they connect with veins rather than the kidney tubules. A thin-walled urinary bladder opens into the amphibian cloaca. *It has no connection with the archinephric ducts*

Amniote kidneys. In reptiles, birds, and mammals three types of kidneys are usually recognized: the pronephros, mesonephros, and metanephros These appear in succession during embryonic development, but only the metanephros persists in the adult. Mesonephros and metanephros actually represent different levels of the opisthonephros of lower forms, the metanephros being equivalent to the posterior portion

The anteriorly located pronephros appears during very early development, but it soon degenerates and the more posterior mesonephros then develops. The duct of the pronephros, however, persists to become the duct of the mesonephros This is actually the same as the archinephric duct but is usually referred to as the Wolffian duct

The mesonephros persists for a time and then degenerates. In the meantime the metanephros has begun to develop from the region posterior to the mesonephros. A few mesonephric tubules and the Wolffian duct persist to contribute to the reproductive system of the male or to remain as vestigial structures

Mesonephros. The tubules of the mesonephros develop in the same manner as opisthonephric tubules. Some of the anterior tubules may even form peritoneal connections. In some forms the mesonephric kidneys become voluminous structures; in others they amount to very little. In reptiles spiny anteaters, and marsupials the mesonephros may even persist for a time after birth.

Metanephros A metanephric kidney develops on each side posterior to the mesonephros It is composed of essentially the same parts as the mesonephros and contains renal corpuscles, secretory tubules, and collecting tubules No peritoneal connections are present Each kidney has a twofold origin. An outgrowth from the posterior end of the Wolffian duct grows forward into the tissue posterior to that from which the mesonephros was derived and which is called the metanephric blastema. The outgrowth is destined to form the ureter and the collecting portion of the kidney It branches and rebranches many times to form ultimately large numbers of fine collecting tubules. At least in mammals at the point where the outgrowth undergoes its primary divisions an expanded region forms the pelvis of the kidney.

From the blastema adjacent to the collecting tubules arise secretory tubules. Each tubule grows [illegible] S-shaped One end establishes connec- [illegible] other expands [illegible] glomerular tuft so that a typical renal corpuscle is formed Each tubule differentiates into several regions (Fig 15). It has been estimated that there are about 1,000,000 renal corpuscles in each human kidney.

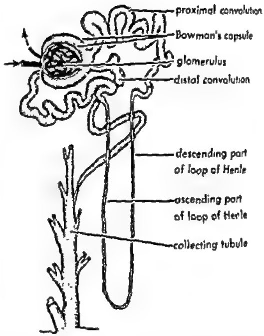


Fig 15 A mammalian metanephric tubule, showing the renal corpuscle and secretory and collecting portions (From C. K Weichert, Anatomy of the Chordates, 2d ed., McGraw-Hill, 1958)

Reptiles. The metanephric kidneys of reptiles lie in the posterior part of the abdominal cavity, usually in the pelvic region They are small, compact, and often markedly lobulated. The posterior portion on each side is somewhat narrower In some lizards the hind parts may even fuse The degree of symmetry varies, being most divergent in snakes and limbless lizards which have notably long, narrow, lobulated kidneys in correlation with the shape of the body One kidney may be entirely behind the other

Snakes and crocodilians lack a urinary bladder but most lizards and turtles have well-developed bladders which open into the cloaca Except in turtles the ureters open independently into the cloaca. In turtles they connect to the bladder. Some turtles possess a pair of accessory urinary bladders which open into the cloaca They may be used as accessory organs of respiration In females they are reported to be filled with water which is used to soften the soil when a nest is being prepared

Birds The kidneys of birds are situated in the pelvic region of the body cavity, their posterior ends are usually joined. They are lobulated structures with short ureters which open independently into the cloaca

Except for the ostrich, birds lack urinary bladders Urinary wastes, chiefly in the form of semisolid uric acid, are eliminated through the cloaca along with the feces

Mammals A rather typical mammalian metanephric kidney (Fig 16) is a compact, bean-shaped organ attached to the dorsal body wall outside the peritoneum. The ureter leaves the medial side at a depression, the hilum At this point a renal vein also leaves the kidney and a renal artery and nerves enter it. The kidney is surrounded by a

against a concentration gradient across the tubular wall. Such a process requires the expenditure of energy. Others have diffused out. Some substances that are reabsorbed are glucose, fructose, amino acids, vitamin C, water, sodium, potassium, calcium, chloride, bicarbonate, sulfate, and phosphate.

The amount of substance that the tubules can re- [illegible]

reabsorption of some substances is regulated to the need of the body and is under hormonal control

Tubular excretion. This phenomenon is the exact antithesis of tubular reabsorption. Material is added to the tubular content through active transport into the tubules Among substances excreted are both normal nitrogenous metabolites, mono- and divalent ions, and several organic compounds. The normal metabolites that undergo tubular excretion in various animal groups include all the nitrogenous waste products, ammonia, urea, trimethylamine oxide, uric acid, creatine, and creatinine. However, the same substance is not always treated in the same manner in different groups of animals. As an example, urea is actively reabsorbed by the renal tubules in the shark and actively excreted by the frog and probably by many other vertebrates Uric acid is actively excreted in bird and reptile tubules but partly reabsorbed in the mammalian tubule.

Many divalent ions, including magnesium, sulfate, thiosulfate, and phosphate, are excreted by the aglomerular fish. Some of these are reabsorbed by the mammalian kidney but it is quite likely that they may be both reabsorbed and excreted in the same kidney Certain foreign substances are found always to be excreted by the renal tubules in all vertebrates and in invertebrates as well Among these are phenol red, diodrast, and para-aminohippuric acid

Tubular excretion is limited by a maximum transfer rate T_{max} similar to the tubular reabsorption. When two substances that are both excreted actively are present at the same time, the T_{max} for one or both may be depressed. This phenomenon is interpreted to mean that the two substances are competing for the same cellular transport system Several substances of widely different chemical nature apparently compete.

Renal clearance tests The glomerular filtration rate (GFR) can be measured indirectly by measuring the excretion of a glomerular substance. A glomerular substance is a substance that is completely filterable but is neither reabsorbed nor actively excreted by the tubules. The amount that is excreted in the urine must therefore equal the amount that was filtered in the same time interval, thus

$$\mathrm{GFR} \times P = U \times V \quad \text{or} \quad \mathrm{GFR} = U \times \frac{V}{P}$$

where P = plasma concentration of the substance, U = concentration in urine, and V = volume formed per unit time. Some glomerular substances employed are inulin, creatinine, mannitol, and ferrocyanide. Of these, the polyfructoside, inulin, appears to be the most reliable in all animals. It is completely filterable and it has neither been found to be reabsorbed nor excreted by the tubules of any animals; it does not appear in the urine of the aglomerular fish. Creatinine is excreted by the tubules of the aglomerular fish and to some extent by many mammals.

By the renal clearance C of any substance is meant the volume of blood that is *completely cleared of the substance per unit time. It is determined similarly to GFR.*

$$C = \frac{U \times V}{P}$$

U and P represent the urine and plasma concentrations of the substance in question and V is urine volume.

The renal clearances of substances that are reabsorbed or excreted are *lower or higher, respectively, than the GFR The amount of a substance T* that is reabsorbed by the *tubules* can be calculated if the GFR and the concentration in plasma P and in urine U of the substance are determined.

$$T = \mathrm{GFR} \times P - UV$$

The net amount of a substance that is actively excreted can be calculated from the *equation*

$$T = UV - \mathrm{GFR} \times P$$

Tubular maximum T_{max} can be measured by determining T at progressively higher plasma concentrations

Renal plasma flow The renal plasma flow (RPF) is the total amount of plasma passing through the kidneys per unit *time In mammals* the kidneys have only an arterial supply of blood, and all the blood that reaches the capillary net around the tubules has passed through the glomeruli first. In all other vertebrates the kidney is supplied with both venous and arterial blood, blood from the renal portal vein goes directly to the capillary net around the tubules.

The RPF can be measured indirectly by measuring the clearance of a substance that is completely excreted by the renal tubules, such as para-aminohippuric acid (PAH), at low plasma levels so that T_{max} is not reached. If all the plasma that passes through the kidneys were cleared of PAH, the PAH clearance would be equal to the RPF. However, measurements of the concentration of PAH in the renal artery and vein have shown that only about 90% of the plasma is cleared of PAH as it passes through the kidney. It is [illegible]

[illegible] substances is called the effective renal plasma flow.

When both RPF and GFR are determined, the fraction of the plasma that is filtered in the glom-

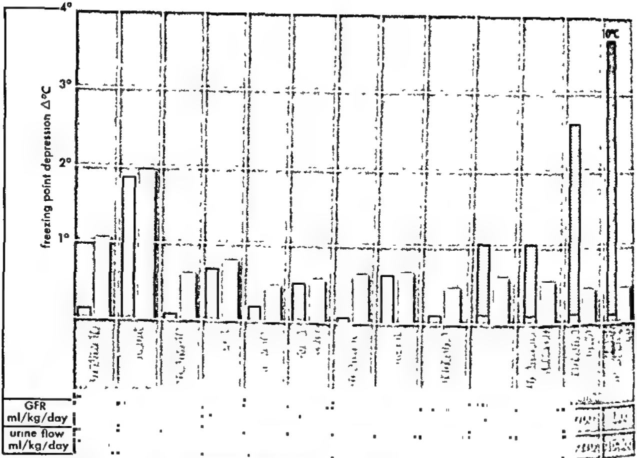


Fig 17. Urea and plasma concentration, glomerular filtration rate and urine flow in various groups of vertebrates. The freezing-point depression in blood is represented by cross-hatched bars, in urine by white bars. The values for glomerular filtration rate (GFR) and urine flow are given in milliliters per kilogram of body weight per day. The difference in values of GFR in the three mammals is a result of the differences in body size. The filtration rate per kilogram body weight decreases with increasing body size. Values given are from various sources. One should not attach too much significance to the individual values but rather to the over-all trend and order of magnitude.

stream while other substances are added from the blood to the tubular content or excreted. Cushny assumed that only reabsorption took place in the tubuli. However, ample evidence that many substances are actively excreted into the tubules was brought forward by E. Marshall (1924) and other investigators.

Marshall also proposed that the same substance may be both excreted and reabsorbed during the passage of the filtrate through the tubule. This important notion has been proved true B. Hargitay and W. Kuhn (1951) proposed that the kidneys of birds and mammals operate as countercurrent multiplier systems. The anatomical arrangement of the nephrons and blood capillaries in parallel loops in mammalian kidneys makes them well suited to operate in this manner. This hypothesis is in agreement with the fact that only birds and mammals have Henle's loops, and only these can produce a urine with a higher osmotic pressure than the blood.

Glomerular activity. The hydrostatic pressure of the blood, derived from the pumping action of the heart, is the driving force that causes plasma to be filtered through the glomerular capillary wall out into Bowman's capsule. The glomerular capillary walls are permeable to all smaller molecules, of molecular weight < 50,000. The filtrate formed is thus free of blood cells and practically protein free With respect to all smaller molecules the composition of the glomerular filtrate is identical to that of the plasma The filtration rate varies in different animal groups according to their habitat (Fig 17). In marine fish, it is very low compared to fresh-water species Some marine fish have no glomeruli whatsoever, and consequently no filtration In others, the number of functioning glomeruli is substantially reduced. In birds and particularly in mammals, the filtration rate is much higher than in any of the lower vertebrates. Regulation of the filtration rate in the individual animal is accomplished through changes in the tonus of the sphincters around the afferent and the efferent glomerular arterioles. In some animals such as the frog, intermittency of the glomerular function has been observed

Tubular reabsorption. Many of the substances that are filtered in the glomeruli are found in much smaller quantities in the final urine. These substances have returned to the blood stream through the tubular wall. Some have been actively reabsorbed, that is, they have transported

during reduced blood flow is known as renin. It is a proteolytic enzyme which acts on a plasma globulin, hypertensinogen, resulting in the formation of a vasoconstrictor substance, hypertensin or angiotonin, now called angiotensin.

Angiotensin This substance has been found in the circulating arterial blood in dogs with experimental renal hypertension and in human beings with essential hypertension Two different forms of the material have been obtained in purified form and their chemical constitution determined The initial product of the action of renin on its substrate is a decapeptide, hypertensin I. This material is inactive. It is transformed into the active pressor substance, hypertensin II or angiotensin, an octapeptide, by a plasma enzyme Angiotensin is destroyed by the enzyme angiotensinase found in blood and tissues

Antirenin When renin is repeatedly injected into the blood stream of an animal antibodies called antirenin are formed Antirenin neutralizes renin and thus decreases the blood pressure in hypertensive animals. In chronic renal hypertension, antirenin formation can be enhanced by repeated injections of renin

Micturition. In nonmammalian vertebrates the anatomical arrangement and origin of urinary ducts and bladder, if present, are completely different from that in mammals. The situation shows extreme variations ranging from direct entrance of kidney

animals

In mammals the urine passes from the renal pelvis through the ureter to the bladder. The urine enters the bladder in a series of squirts, about 2–6 per minute, caused by rhythmic peristaltic movements of the ureter wall An oblique valvular opening prevents the backflow of urine from bladder to ureter

Retention of urine During the gradual accumulation of urine, the pressure in the bladder remains relatively constant at about 5 centimeters (cm) H_2O This is a result of the adjustment of the detrusor muscle in the bladder wall When the filling reaches a certain value in adult man about 200–400 milliliters (ml) further filling brings about a steeper rise in pressure When the pressure reaches about 20-30 cm H_2O the urge to micturate arises It can be voluntarily overruled until about 700 ml have accumulated

Reflex mechanism The bladder is innervated by parasympathetic and sympathetic nerves Stimulation of the parasympathetic nerves causes contractions of the bladder Sympathetic stimulation causes relaxation of the detrusor muscle and contraction of the internal sphincter of the urethra.

A rise in bladder pressure acts on the bladder wall, causing increasing rhythmic contractions Furthermore, the stretching of the wall gives impulses in sensory nerve endings These impulses travel up to the nervous center in the upper lumbar region and are summated until, when released, they result in prolonged intensive contraction of the musculature of the bladder, opening of external and internal sphincters, straightening of the urethra, and relaxation of the peritoneum. Voluntary restraint of micturition involves chiefly inhibition of spinal reflexes, which results in relaxation of the detrusor muscle, abolition of rhythmic contractions, and closure of the internal sphincter.

[B S N.]

Bibliography: C. W. Gottschalk and M. Mylle, Evidence that the mammalian nephron functions as a countercurrent multiplier system, *Science*, 128: 594, 1958; B Hargitay and W. Kuhn, Das Multiplikationsprinzip als Grundlage der Harnkonzentrierung in der Niere, *Z. Elektrochem.*, 55:539, 1951; E K. Marshall, The comparative physiology of the kidney in relation to theories of renal secretion, *Physiol. Revs.*, 14:133-159, 1934; B. Schmidt-Nielsen...

Urination

The process by which certain soluble materials in water solution are evacuated from the urinary bladder and thus expelled from the body. Exactly what quantity of various materials are passed into the bladder to be urinated is under the control of the kidneys The primary function of this excretory process is twofold: (1) the excretion of certain waste materials accumulated in the body as a result of normal metabolic processes, and (2) the discharge of certain solutes, along with body water, in the interest of maintaining the proper levels of these factors in the internal environment of the body.

In man, the act of urination involves both reflex and voluntary components Although under normal circumstances the final initiation of urination is voluntary, such control is not possible until the reflex components of the process are appropriately triggered Urination can be voluntarily terminated, however, at any point in the execution of the act. *Micturition*, as this excretory process is also called, is under the combined regulation of the ...

...ounted to the central nervous system via the pelvic nerve

While the spinal cord alone can mediate reflex urination, higher levels of the nervous system also exert urinary control in the normal individual. Facilitory centers exist at the level of the pons and posterior ...

...ilitory centers have been localized, in the orbital frontal area and the cingulate gyrus See BRAIN.

eruli (FF) can be determined: FF = GFR/RPF. In man, FF is approximately 20%. It varies from about 10-50% in different mammals. The FF is much lower in vertebrates with a renal portal system than in mammals.

Water and salt in the mammalian tubule. During the passage of the glomerular filtrate through the proximal convoluted tubule, approximately seven-eighths of the water and salt are reabsorbed The tubular fluid remains isotonic to the plasma throughout the proximal tubule. As the fluid proceeds down the loop of Henle, it becomes increasingly concentrated, presumably because the loop acts as a countercurrent multiplier system As the fluid proceeds up the ascending limb of the loop of Henle, it again becomes increasingly more dilute because of reabsorption of sodium, and at the point where the convoluted tubule begins, the fluid tonicity is only one-third of that of the plasma.

Presumably, the permeability to water of the distal convoluted tubule and collecting duct is regulated by antidiuretic hormone (ADH) from the posterior pituitary. In the absence of ADH, the permeability to water is low and the tubular fluid remains hypotonic to the blood during its passage through the tubule. The still hypotonic fluid passes down through the collecting ducts, and thus the urine is also hypotonic to the plasma. *See* PITUITARY GLAND.

In the presence of ADH, the distal tubule and the collecting ducts become permeable to water. Water diffuses out and the fluid becomes isotonic to the surrounding tissue and blood. As the fluid enters the collecting duct and proceeds down through the outer and inner zone of the medulla it passes through tissue with increasing osmotic concentration. Water diffuses out of the collecting ducts and the fluid inside the collecting ducts becomes increasingly concentrated. In this manner, a hypertonic urine is produced

Diuresis. The kidneys respond to excess intake of water by excreting an osmotically dilute urine water diuresis. The maximum possible urine flow in man during water diuresis corresponds to ⅛–1/10 of the GFR, or 15-20 liters in 24 hours. In this state of diuresis, proximal reabsorption of water is normal while distal water reabsorption is minimal The antidiuretic hormone is presumably absent from the blood stream

When large amounts of solutes, such as sodium chloride, urea, mannitol, or sucrose, are filtered in the glomeruli, the proximal reabsorption of water is impaired because of unreabsorbed solute present in the tubular fluid (osmotic diuresis). In mammals, the urine is hypertonic to the blood during osmotic diuresis; in lower vertebrates, it is slightly hypotonic. Osmotic diuresis occurs in teleosts, both glomerular and aglomerular, when handled or injured. Na^+ and Cl^-, normally almost absent from the urine, are then excreted in large quantities. Because proximal water reabsorption is impaired during osmotic diuresis, the maximal diuresis can reach values almost as high as the filtration rate. A number of substances that act as osmotic diuretics and substances that interfere with the tubular function in other ways are used to produce diuresis All soluble mercury compounds act directly on the proximal tubule, decreasing the reabsorption of sodium and water Organic mercury compounds of relatively low toxicity are used clinically.

Acid-base equilibrium. The hydrogen ion concentration of the plasma is closely regulated in most animals. In aquatic vertebrates, the pH of the plasma is between 7.2 and 7.8, regardless of medium pH. In terrestrial vertebrates, the blood pH is usually near 7.4 The constancy of the pH is the result of three factors: the excellent buffer capacity of the blood, the respiratory regulation of the CO_2 tension, and the renal regulation of acid and base excretion.

The carbonic acid–bicarbonate system, H_2CO_3–HCO_3^-, is the most important buffer system in the blood. The amount of HCO_3^- present is determined by the amount of cations, such as Na^+, K^+, Ca^{++} and others, in excess of other anions such as Cl^-, SO_4^{--} and $H_2PO_4^-$. This excess of cations is called the alkali reserve. The amount of H_2CO_3 is determined by the partial pressure of CO_2. The renal regulation is concerned with the maintenance of the alkali reserve. The renal tubules compensate for a lowering in the plasma pH by reabsorbing Na^+ in exchange for H^+. This acidifies the tubular fluid, but the urinary pH does not decrease below 4.8. Further acidification of the urine is prevented by the addition of ammonia. The tubular cells contain the enzyme glutaminase which releases ammonia from the amino acid glutamine. Ammonia diffuses into the tubules and is trapped there as the ammonium ion, NH_4^+.

Renal compensation for a rise in the plasma pH is accomplished through the excretion of Na^+ and K^+ cations in combination with HCO_3^- or HPO_4^{--}. The pH of the urine may become as high as 8.4

In the frog the acidification of the urine has been found to take place in the distal tubule In mammals, the acidification is localized in the distal part of the distal tubule or in the collecting ducts. Ammonia secretion, potassium secretion, and sodium reabsorption occur in the same place It is generally believed that sodium reabsorption in this segment always takes place through an ion-exchange mechanism, that is, tubular sodium is exchanged with either potassium or hydrogen ions

Hypertension of renal origin. When the renal artery in an animal is compressed experimentally, a substance is released from the kidney into the blood which causes hypertension This experimental hypertension resembles essential hypertension in man, in which preglomerular arteriolar sclerosis and reduction in renal blood flow is found. It has been stated by some investigators that all hypertension is of renal origin, but there is no general agreement

Renin The substance released from the kidney

Uropygi

An order of arachnids, the tailed whip scorpions, comprising about 70 species from tropical and warm temperate Asia and the Americas. Most are dark reddish-brown, of medium to giant size, 18–65 mm, the largest one being *Mastigoproctus giganteus* of the southern United States and Mexico. The elongate, flattened body bears in front a pair of greatly thickened, raptorial pedipalps set with many sharp spines and used to hold and crush insect prey. The first pair of legs is elongated and modified into feelers. The abdomen terminates in a slender, many jointed, whiplike flagellum. The uropygids are harmless, nocturnal creatures without poison glands that live in dark places and burrow into the soil. When disturbed, they expel a volatile liquid, with the strong odor of acetic acid, from a gland at the base of the tail. This accounts for the name "vinegaroon" given by many Americans to these much-feared animals. See ARACHNIDA. [W.J.GH.]

Uropygial gland

The only skin gland possessed by most birds. It is a large, compact, bilobed secretory organ located at the base of the tail, the uropygium, and is present in most carinate birds in which the sternum has a ridge or keel. It varies widely in size, shape, and structure among different species. It is large in most aquatic birds, rudimentary in many nighthawks and some pigeons, absent in bustards, many pigeons and parrots, and in all ratite birds (ostrich, cassowary, emu, and others) although it is present in the embryos of all. The secretion, predominantly oily and sometimes of offensive odor as in the musk-duck, hoopoe, and petrel, is discharged through an orifice at the tip of a nipplelike protuberance which is often encircled by short, bristly feathers. Its use as an anointment for rendering feath[illegible]
gera [illegible]
beer [illegible]
uses [illegible] [M.L.R.]

Ursa Major

The most widely known and oldest of the astronomical constellations. Ursa Major or the Great Bear is a circumpolar group as viewed from the middle latitudes of the Northern Hemisphere. One part of the configuration, a group of seven bright stars, which is pictured as the tail of the Great Bear, is commonly known in the United States as the Big Dipper which it resembles. This group of stars is also known in various lands as Charles' Wain (wagon) and the Plough. The Chinese call it the Northern Bushel. The two stars at the front of the bowl of the dipper are called pointers, because a line joining them points to Polaris, the North Star. One of the pointers, the northern one, called Dubhe, is a navigational star. The star next to the end of the handle is Mizar, another navigational star, with its close companion Alcor. Next to Mizar is Alioth, the third navigational star in this group. See CONSTELLATION. [C.S.Y.]

Ursa Minor

The astronomical constellation Little Bear. Ursa Minor is a circumpolar constellation whose brightest star, Polaris, is almost at the North Celestial Pole. Seven of the eight stars appear to form a dipper, hence the constellation is alternately known as the Little Dipper. Polaris is at the end of the handle. Situated about 1° from the true celestial pole, Polaris is a variable star, pulsating in brightness periodically. The two bright stars at the front of the bowl are often called the Guardian of the Pole, because they circle about Polaris closer than other conspicuous stars. Kochab, the brighter of the two, was at one time closer to the true Pole. It is an Arab word meaning polestar. The Little Bear is also known as Smaller Chariot by the Danes. See CONSTELLATION; URSA MAJOR. [C.S.Y.]

Urticales

An order of the plant subclass Dicotyledoneae including 4 families with 32 genera and about 1750 species. These plants range from low, delicate herbs through coarse, tall herbs and shrubs to tall trees. The elm family (Ulmaceae) with 15 genera and 150 species, mostly in the tropical and temperate zones, includes the elms and the hackberry. The mulberry family (Moraceae) has 73 genera and 1000 species, mostly tropical and subtropical. The breadfruit and multiple fruits of the red mulberry are edible. Leaves of the white mulberry are the food of silk worms. The wood of osage-orange (*Maclura*) is preferred for making bows, and the fiber of the paper mulberry (*Broussonetia*) is used in making paper and tapa cloth. The nettle family (Urticaceae) with 42 genera and 600 species, mostly tropical and subtropical, is of little economic importance. The hemp family (Cannabinaceae) has but 2 genera and 3 species. Hemp (*Cannabis sativa*) produces a valuable fiber and marijuana, a narcotic drug used illegally in cigarettes. Ramie (*Boehmeria nivea*) is the source of a very tough fiber, and hops (*Humulus lupulus*) are used in making beer. See BREADFRUIT; ELM; HACKBERRY; HEMP; HOP; MARIJUANA; MULBERRY; OSAGE-ORANGE; RAMIE; see also DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM; TREE. [P.D.S.]

Ustilaginales

An order of the subclass Heterobasidiomycetidae comprising the smut fungi which parasitize plants —chiefly members of the grass and sedge families —and cause diseases known as smut or bunt. In Ustilaginales, intercalary cells of the binucleate mycelium develop into teleutospores which function as hypobasidia. There are approximately 37 genera and 700 species divided into 3 families: the Ustilaginaceae in which basidiospores bud from the sides of the septate epibasidium, the Til-

That the act of urination is more than a simple innate act is shown by the degree of learned control that is possible in higher animals and man. Moreover, animal experiments concerned with the differences in the pattern of male and female urination show that, in part, the details of the urinary act are under the control of the sex hormones. *See* BODY RHYTHM; HORMONE. [R.A.M.]

Urine

An aqueous solution of organic and inorganic substances, mostly waste products of tissue metabolism. These are selectively filtered from the blood by the kidney and then excreted through the ureters to the bladder. Average adult human excretion is 1500 ml daily, of which 95% is water containing about 60 g of solutes. Urea comprises 30 g, inorganic salts 25 g, and the remainder is mostly creatinine, uric acid, and ammonia. Normal urine is yellow, aromatic, and slightly acid. The specific gravity varies from 1.010 to 1.030. Quantity and content are greatly affected by fluid intake and unusual loss, as for example, through perspiration, vomiting, diarrhea, or hemorrhage. *See* URINARY SYSTEM. [E.C.ST.]

Urogenital system

The combined structures comprising the urinary and genital, or reproductive, organs of vertebrates. The terms urinogenital or genitourinary system are equally applicable and just as frequently used. During embryonic development, the urinary and reproductive systems are closely interrelated, as their ducts arise in associated mesodermal regions. Common passages (Fig. 1) are associated with both systems in various vertebrates, such as the single orifice for the emission of both urine and sperm.

Urinary system. The principal organ of the urinary system is the kidney with its associated structures. Three morphologically distinct kidneys occur among the vertebrates. These are the pronephros, mesonephros, and metanephros. The pronephros is drained by the pronephric duct, the mesonephros by the Wolffian duct, and the metanephros by the ureter. The principal function of the urinary system is the removal of nitrogenous waste products. *See* KIDNEY; URINARY SYSTEM.

Reproductive system. While the same structures of the urinary system occur in both males and females, specific structures, such as the ovaries and testes, are found in the normal female and male, respectively. The primary function of the reproductive system is the perpetuation of the species. The main reproductive structures of the female vertebrate consist of the ovary, oviduct, uterus, vagina, and external orifice. Many modifications occur among these structures (Fig. 2) in the various classes. Essentially, the male system has the testes and their ducts, seminal vesicles, prostate gland, Cowper's gland, and penis. Modifications of these structures also occur among the various vertebrate species. *See* REPRODUCTIVE SYSTEM. [C.B.C.]

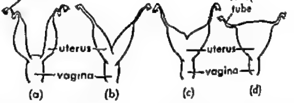


Fig. 2. Four types of mammalian uteri. (*a*) Duplex. (*b*) Bipartite. (*c*) Bicornuate. (*d*) Simplex.

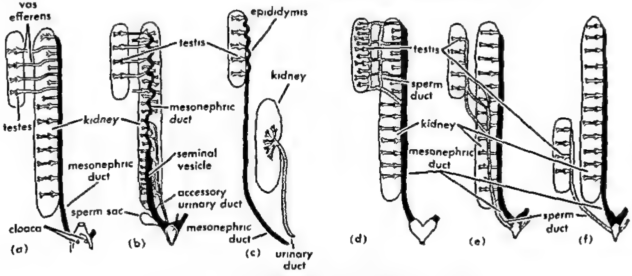
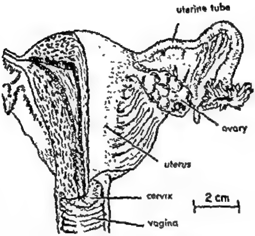


Fig. 1. Mesonephric duct of male urogenital system. (*a*) Basic plan, both urine and sperm are carried. (*b*) Dogfish, sperm and some urine carried. (*c*) Amniote, sperm only. (*d*, *e*, *f*) Sturgeon, lungfish, and teleost showing transition toward a sperm duct independent of the mesonephric duct.



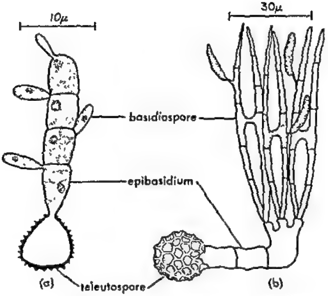
Uterus (From L B Arey, *Developmental Anatomy*, 6th ed, Saunders, 1954)

projects into the vagina Normally the uterus is tilted slightly forward and lies behind the urinary bladder.

The lining, or mucosa, responds to hormonal stimulation to produce the changes of the menstrual cycle. If fertilization does not occur, the thickened, vascular lining is sloughed off and a new cycle begins. When pregnancy occurs the mucosa continues to thicken to form an intimate connection with the developing placenta However, the maternal and fetal bloods do not mix.

The uterus develops from two embryonic tubes, which may remain distinct in the adult, forming two uterine chambers as in the rabbit; or may fuse to various degrees, forming a branched two-chambered uterus with a single lower portion, as in the cow; or may unite to form a single chamber, as in the human species. *See* REPRODUCTIVE SYSTEM

[E.G.ST.]



Germinating teleutospores of Ustilaginales (a) *Ustilago scabiosae* (after R A Harper, *Wisconsin Acad of Sciences Transaction*, 12 475–488, 1898) (b) *Tilletia caries* (after G. M. Smith, *Cryptogamic Botany*, vol 1, 2d ed., McGraw-Hill, 1955).

letiaceae in which basidiospores form at the tip of the *epibasidium*, and the Graphiolaceae in which teleutospores are produced in a cuplike *fruiting* body. *See* HETEROBASIDIOMYCETIDAE, SMUT (MICROBIOLOGY) [R.M.P.]

Bibliography: C. J Alexopoulos, *Introductory Mycology*, 1952; G. M. Smith, *Cryptogamic Botany*, vol. 1, 2d ed., 1955.

Uterine disorders

Uterine and cervical disorders account for a large portion of medical practice Although the cervix is not considered a part of the uterus, cervical disorders are included in this article because of the close proximity of the two areas. The major categories include congenital, inflammatory, neoplastic, and functional diseases *See* UTERUS.

Failure of formation or developmental abnormalities are not uncommon, but most aberrations are of minor significance, such as a double uterus. Infantile uteri may result from lack of hormonal stimulation and are found in a number of endocrine disturbances. *See* HORMONE.

Inflammations of the uterus are not common, but when present, follow infections resulting from trauma, childbirth, or miscarriage. However, cervicitis, an inflammation of the uterine cervix, is probably the most frequently encountered female disease Most cervicitis is nonspecific and tends to become chronic, but specific types are caused by staphylococci, streptococci, gonococci, syphilis, chancroid, and tuberculosis. *See* GONORRHEA; SOFT CHANCRE; STAPHYLOCOCCUS; STREPTOCOCCUS; SYPHILIS; TUBERCULOSIS.

The uterus is frequently the site of both benign and malignant tumor formation and the three most common growths are as follows (1) Endometrial polyps are benign, simple or multiple masses which are often sources of abnormal bleeding; they are otherwise unimportant (2) Leiomyomas, or fibroids, are the most common tumors in women, and occur in 10% of women of reproductive age. These are benign smooth-muscle growths, varying in size from minute bodies to huge masses which fill the pelvis and lower abdomen They may be asymptomatic or may cause abnormal bleeding The larger fibroids commonly produce signs or symptoms as a result of the mechanical effects on adjacent organs They may also interfere with pregnancy, labor, or delivery. (3) Carcinoma of the endometrium is an important female malignancy but has a lower incidence than breast or cervical cancer. It is most often found in the 50- to 70-year age group and in patients with obesity, hypertension, and diabetes mellitus. Unlike cervical cancer, endometrial carcinoma occurs most often in childless women. Prognosis is good if diagnosis is early

Cervical carcinomas are the most common malignancies of the female genital system, and occur predominately in the 35- to 60-year age group. There is an increased incidence with successive pregnancies; they are not common in virgins, Jews, and Moslems. In the United States, cervical cancer occurs twice as often in nonwhite as in white women. Early diagnosis, with correspondingly higher rates of cure, has been greatly advanced by publicity, the development of cytologic examination of vaginal smears, and the increasing use of routine cervical biopsies in other cervical disorders

Polyps of the cervical canal are almost always benign and occur in 3–5% of all women. They are frequently the source of irregular bleeding or "spotting" but are otherwise clinically unimportant. *See* ONCOLOGY

Functional disorders of the uterine mucosa produce menstrual irregularities and other symptoms and, for the most part, stem from obscure causes The complex endocrine regulation of the very reactive endometrium is subject to a great many other individual factors. *Hyperplasias*, atrophies, and displaced endometrial tissue are seen frequently The latter, known as endometriosis, is marked by bits of mucosa located in the pelvis, ovaries, appendix, or even the nasal lining These cell groups may react to cyclic hormonal stimulation, thereby causing sometimes bizarre symptoms and internal bleeding. The more common hyperplasias and atrophies account for about one-fifth of all gynecologic disorders and are the most frequent cause of abnormal bleeding *See* ENDOCRINE SYSTEM; MENSTRUATION [E C ST.]

Uterus

The hollow, muscular womb. An adult human uterus, before pregnancy, measures 3 by 2 by 1 in. in size and has the shape of an inverted, flattened pear. It receives the Fallopian tubes at its upper corners; the lower, narrowed tion, the cervix,

V Vaccination to Vulture

Vaccination

Originally, the active immunization against smallpox by injection of the cowpox virus, vaccinia. The term has come to include many techniques of immunization, both active and passive, in which a vaccine or biological is employed. Active immunization refers to the production of antibody in the human or animal following injection of a specific antibody inducer, termed an antigen; passive immunization refers to the transfer to one individual of antibody already formed in another. In many cases illness itself produces antibody formation and is natural, active immunization. *See* ANTIBODY; ANTIGEN; BIOLOGICALS; IMMUNITY.

Vaccines are available against diphtheria, whooping cough, tetanus, typhoid fever, measles, mumps, rabies, poliomyelitis, cholera, typhus, and certain other diseases. Vaccination may be used for prophylaxis to protect against the disease, or to modify its course favorably. A vaccine must be safe, effective, and easy to administer. The duration of effectiveness varies greatly with the individual, age, and the specific disease. See also articles on individual diseases. [E.G.ST.]

Vacuum cleaner

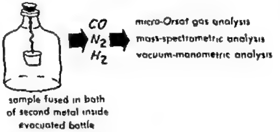
A mechanical appliance for the dry removal of dust and loose dirt from fabrics and surfaces. Vacuum cleaners are widely used in domestic and industrial cleaning of surfaces that cannot be wiped or brushed, such as carpets, upholstery, tapestry, or highly contoured surfaces. In operation, atmospheric pressure forces air into the cleaning tool, the air carrying with it loose dust and dirt particles. The dust-laden air then usually passes through a filter bag that retains the dust. The air is exhausted through an electric-driven fan which provides the vacuum or low internal pressure that enables atmospheric air to force the dust into the machine. A vacuum cleaner has the great advantage that the dirt is retained within the filter bag. However, the force of the draft is limited by the use of atmospheric pressure. For this reason, industrial cleaning of work in process is often by air blast, with which larger and more adherent particles can be removed. [F.H.R.]

Vacuum fusion

A technique of analytical chemistry for determining the oxygen, hydrogen, and sometimes nitrogen content of metals. The method can be applied to a wide variety of metals, the alkali and alkaline earth metals being exceptions. The range of the method extends from 1% down to a few parts per million for oxygen and nitrogen and down to fractional parts per million for hydrogen.

The metal sample is either fused or dissolved in a bath, or flux, of a second metal in a heated graphite crucible supported inside an evacuated glass or quartz vessel. Oxygen is released from the metal as carbon monoxide by reaction of oxides or dissolved oxygen with carbon from the graphite crucible at high temperature. Metal nitrides dissociate, although not always quantitatively, to form elemental nitrogen. Hydrogen is evolved as elemental hydrogen.

The mixture of carbon monoxide (CO), nitrogen (N_2), and hydrogen (H_2) is analyzed to determine individual component concentrations by one of several techniques, including micro-Orsat, mass-spectrometric, and vacuum-manometric procedures. The latter is in most general use. Gas quantities



Vacuum fusion methods of analysis

are determined by measurement of the pressure of the gas after confinement into a small, calibrated volume. The product of the pressure and volume is proportional to the total number of moles of gas present, regardless of the species of gas molecules. Measurement is first made of the total quantity of CO, N_2, and H_2 collected. These gases are then oxidized to carbon dioxide (CO_2), water (H_2O), and N_2 by passing over hot copper oxide. The CO_2 and H_2O can be removed from the gas mixture successively by selective freezing or by use of chemical absorbents. The decrease in total moles of gas accompanying removal of each species permits determination of the number of moles of CO_2, N_2, and H_2O.

Furnace temperatures and bath conditions must be selected for each metal or alloy to ensure quantitative recovery of oxygen. The required furnace temperatures range from 1650°C for iron, nickel, and low alloy steels to as high as 1900°C for metals

als in retorts and fused salt electrolytic processes, although inert gas or hydrogen atmospheres are actually employed. Reduced pressure has a direct effect in carbon-reduction processes, for example, in the reduction of magnesium and niobium oxides and the manufacture of low carbon ferrochrome. Hot ire, iodide decomposition processes, used for production of especially pure titanium, zirconium and ther metals are carried out under reduced pressure. The Mond process for high-purity nickel is a imilar process.

Vacuum distillation processes include vacuum plating, such as the deposition of vaporized metals, such as aluminum, on other metals, plastics, and glass. Vacuum distillation is also used for refining, for example, in the removal of zinc from lead.

Analytical techniques for gases in metals involve gas extraction under vacuum from both liquid and solid *See* METAL COATINGS, METALLURGY, POWDER METALLURGY, PYROMETALLURGY, VACUUM FUSION. [T B K]

Bibliography: R Bunshah (ed), *Vacuum Metallurgy*, 1958.

Vacuum pump

The vacuum pump differs from other pumps because the pressure driving a gas into the pump is less than atmospheric and becomes vanishingly small at high vacuums; therefore the entrance must be large and free from obstruction The pump operates over a much wider range than its compressor counterpart, a span of 10^9 from atmosphere to 10^{-8} mm Hg being commonplace To cover the whole range two or more stages are used. The first, discharging to the air, is a backing pump and the last a high-vacuum or fine pump, intermediate pumps are called boosters The table gives average ranges and Fig 1 shows working characteristics

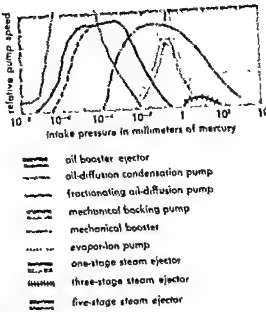


Fig 1 Working characteristics of vacuum pumps; if pump speed is in liters per second, ordinate extends from 1 to 10,000 in logarithmic steps

Backing pumps. There are two principal types of backing pumps: mechanical, and steam ejector. The work horse of vacuum practice is the single- or double-stage rotary-displacement mechanical pump mounted on baseboard with electric motor; typical construction of the pump portion is shown in Fig 2 Because of the reexpansion of any gas left

Representative vacuum pump ranges and characteristics

Description of pump	Capacity	Pressure, mm Hg: Exhaust	Pressure, mm Hg: Intake	Usual or trade names
Backing pumps				
Oil-sealed, mechanical				Duoseal, Hyvac, Kinney, Metrovac, Stokes
1-stage	0 1–100 cfm	atm	0.1–0 01	
2-stage	0.1–100 cfm	atm	0 001	
Steam ejectors				
1-stage	10 cfm	atm	100–50	
2-stage	100 cfm	atm	20–10	
3-stage	1000 cfm	atm	5–0 5	
4-stage	10 000 cfm	atm	0 1–0 05	
Boosters				
Mechanical Roots-Connersville type	10–1000 cfm	5–1	0 3–0.03	Windmill pumps
Oil vapor ejectors	10–1000 cfm	10–3	0 5–0 03	
High vacuum pumps				
Diffusion condensation				
Mercury filling no cold trap	1–10,000 liters/sec	3–0 3	3×10^{-6}	
Mercury filling, liquid air trap	1–10 000 liters/sec	3–0 3	$<10^{-7}$	
Oil filling, no cold trap	1–30,000 liters/sec	0 2–0 05	5×10^{-6}	
Oil filling fractionating	1–30,000 liters/sec	0 2–0 05	10^{-7}	
Oil filling cold trap	1–30,000 liters/sec	0 2–0 05	5×10^{-8}	
Getter pumps	1000–10,000 liters/sec	10^{-6}	10^{-6}–10^{-10}	Evaporion

such as titanium, zirconium, tantalum, and thorium Iron, nickel, and low alloy steels are usually fused without addition of other metals. High-melting metals require use of a second metal to lower their melting points. The second metal is used as a previously prepared bath or, in some instances, is added simultaneously with the sample metal. Iron, nickel, platinum, and tin are the most generally used fluxing metals.

Nitrogen is not always quantitatively evolved, and values obtained by vacuum-fusion methods for nitrogen are not generally reliable. Hydrogen is always evolved quantitatively upon fusion of the metal in vacuum. When hydrogen alone is of interest, simpler and more rapid techniques can be used involving vacuum-extraction from metal heated to temperatures well below the melting point. *See* GAS ANALYSIS. [F.C.B]

Vacuum measurement

The determination of a fluid pressure less in magnitude than the pressure of the atmosphere.

These low pressures are perhaps best expressed in terms of the height, in millimeters, of the column of mercury which the given pressure (vacuum) will support, referred to zero pressure The height of the column of mercury which the pressure will support may also be expressed in microns (a micron is one-millionth of a meter) or in inches Less common units of measurement are fractions of an atmosphere and direct measure of force per unit area, such as pounds per square inch

Atmospheric pressure is sometimes used as a reference particularly for relatively crude measurements The pressure of the standard atmosphere is 29.92 in. of mercury; therefore, using atmospheric pressure as a reference, an absolute pressure of 10 in. of mercury would be expressed as 19 9 in of vacuum.

In the laboratory, measurement of vacuum is important because the vacuum (pressure) level has a significant effect on most physical, chemical, and biological processes.

In industry, vacuum level is commonly measured and controlled to maintain uniformity of product and as a guide to safe operation. Vacuum measurement is used, for example, in the lamp industry (evacuation of bulbs); in metallurgy (treatment of metals attacked by common gases), and in the pharmaceutical industry (distilling heat sensitive compounds). *See* PRESSURE CONTROL, AUTOMATIC.

Pressures above 1 millimeter of mercury—under some conditions, even lower pressures—can be measured directly by familiar pressure gages: liquid-column gages, diaphragm-pressure gages, bellows gages, and bourdon-spring gages (*see* BOURDON-SPRING PRESSURE GAGE; MANOMETER; PRESSURE MEASUREMENT) At pressures below about 1 millimeter of mercury, mechanical effects such as hysteresis, ambient errors, and vibration make these gages impracticable.

Pressures below *this level* are best measured by gages which infer the pressure from the measurement of some other property of the gas, such as thermal conductivity or ionization (*see* IONIZATION GAGE; PIRANI GAGE). The thermocouple gage, similar to the Pirani gage, also is widely used. Less frequently used are the Knudsen gage and the rotating viscometer gage.

The McLeod gage is used as an absolute standard of pressure in the range of 0 0001–10 millimeters of mercury (*see* MCLEOD GAGE).

[B D H ; H C P]

Bibliography: D. M. Considine (ed.), *Process Instruments and Controls Handbook*, 1957; S Dushman, *Scientific Foundations of Vacuum Techniques*, 1949.

Vacuum metallurgy

The processing of metals or metallic compounds under reduced pressure to facilitate desirable reactions such as degassing, or to prevent other reactions, such as atmospheric contamination.

Vacuum melting is employed on high-quality steels and high-temperature alloys to prevent atmospheric oxidation, to allow deoxidation by reaction with carbon in the melt, and to remove other dissolved gases, such as hydrogen and nitrogen Significant improvements in mechanical properties are thus obtained. Reactive metals, such as titanium, zirconium, and hafnium, are vacuum melted to avoid contamination by oxygen and nitrogen Many other nonferrous metals are vacuum melted to remove dissolved gases Casting under vacuum may be an integral part of the operation In vacuum induction melting, induction furnaces are enclosed, often with a vacuum lock for admission of casting molds, in a large evacuated chamber.

Contact with refractories is undesirable for reactive metals and is avoided by the following modifications. In the consumable-electrode vacuum arc melting process, the metal is melted in a water cooled copper crucible by an arc struck between the metal pool and a consumable electrode, made of the impure metal The arc may be replaced by an electron beam which enables higher temperatures to be reached In skull-melting, a water-cooled copper crucible which can be tilted for casting and either a consumable or a permanent tungsten electrode are used.

Degassing of liquid metals can also be carried out in processes separate from melting Thus steels, especially for large forgings, are degassed in a ladle or during pouring into a mold enclosed in a vacuum chamber.

Applications of vacuum methods to solids include many heat-treatment operations, carried out under vacuum in retort-type furnaces to prevent atmospheric contamination and to remove dissolved gases Sintering of powdered products, including metals and cermets, and brazing and soldering operations are further examples The electronics industry uses vacuum processing in the manufacture of capacitors and tubes.

Vacuum extraction processes are usually defined to include reduction of the halides of reactive met

longed periods with only occasional auxiliary pumping

Measuring pump speeds. All methods require a vacuum gage. Accurate measurement of speed under highest vacuum requires special apparatus and training. An approximate measurement can be made by comparing atmospheric leakage and pressure differential. Air is admitted through an adjustable leak from a graduated tube which dips into an oil bath. The fine pressure at the intake of the pump and the rate of rise of the oil in the measuring tube are both recorded. Speed S is calculated from

$$S = LP/1000p$$

where S in pumping speed in liters per second, L is leak rate in cubic centimeters per second of the gas at a pressure P, and p is the fine pressure of the pump in the same units as p. For example, if leak rate L is 1 cc/sec of air at an atmospheric pressure P of 750 mm Hg and fine pressure p is 5×10^{-5} mm Hg ...

Vacuum tube

A subclass of electron tubes. This class includes all tubes pumped to a sufficiently high vacuum so that electron motion may exist through the space between the various tube elements with a negligible number of coll...

...tubes and various classifications are possible. A tube manual put out by one of the largest manufacturers consists of four volumes and lists the characteristics of hundreds of tube types. Broadly speaking, vacuum tubes are classified as receiving tubes, transmitting tubes, phototubes, cathode-ray tubes including television picture tubes, and special tubes.

Receiving tubes are low-voltage and low-power tubes which are used in receivers, computers, and sensitive control equipment. Included are such tubes as diodes, triodes, tetrodes, and pentodes (defined in the next section). Transmitting tubes are high-power tubes generally used in radio-frequency transmitting equipment and they are classified by type the same as receiving tubes. Phototubes are light sensitive tubes used in sound film equipment, door openers, and a host of industrial applications. Cathode ray tubes are used as indicator tubes and television tubes. They present a light image in response to an electrical input. Various special tubes include mixers, indicators, and ...

... multigrid tubes will be presented, including some special purpose tubes. For a complete discussion of the following special purpose tubes, see BACKWARD-WAVE TUBE, CATHODE RAY TUBE, KINESCOPE, KLYSTRON, MICROWAVE TUBE; PHOTOTUBE; STORAGE TUBE; TELEVISION CAMERA TUBE; TRAVELING-WAVE TUBE, X-RAY TUBE.

Basic operation. Vacuum tubes depend upon two basic physical phenomena for their operation. The first is the copious emission of electrons by certain elements and compounds when the energy levels of the surface atoms are raised by the addition of heat (thermionic emission), light photons (photoemission), kinetic energy of bombarding particles (secondary emission), or potential energy (high field emission). *See* ELECTRON EMISSION. The second phenomenon is the mechanical force exerted by an electric field upon an electron, thus permitting the control of the movement of large numbers of the electrons within the tube. *See* ELECTRON MOTION IN VACUUM.

Vacuum tubes therefore consist of an electrode capable of electron emission (cathode or filament) and one or more electrodes for establishing and controlling an electric field about the cathode and thereby controlling the movement of the emitted electrons between the cathode and collection electrodes (anodes). Electron tubes are all basically unilateral circuit elements inasmuch as the flow of electrons can only exist from the cathode to an anode. That is, conventional current can only exist from anode to cathode since conventional current flow is assumed opposite to that of actual electron flow.

A diode vacuum tube contains two electrodes, the cathode and an anode to collect the electrons. The anode current is a function of the anode to cathode voltage drop (see DIODE, VACUUM). Anode current may also be controlled by the insertion of

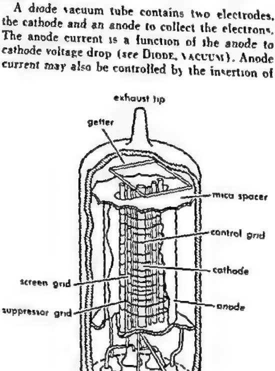


Fig. 1. Typical construction of miniature vacuum tube

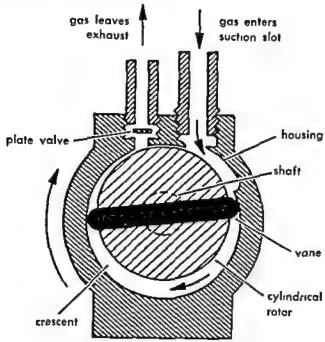


Fig. 2 Typical mechanical backing pump.

after the compression stroke, it is necessary to replace the gas with a noncompressible sealing oil, which is generally admitted through the shaft seal, the pump being immersed in an oil tank The oil eventually becomes contaminated with volatiles and water droplets and must be replaced Gas ballast mechanical pumps continuously purify the oil by a controllable air bleed that blows out the water vapor.

Steam ejectors are stationary pumps (Fig 3) in which the gas is entrained and removed by high-velocity jets of steam They are preferred for large installations, expecially where objectionable vapors should be flushed to waste A three-stage unit generally includes a barometric condenser (*see* CONDENSER, VAPOR) between the second and third stage

Booster pumps. The mechanical booster, patterned after the 2- or 3-lobed gear pump (*see* PUMP) and designed with wide clearances for low friction operation has come into use where large throughputs must be handled, for instance, in vacuum smelting and in the continuous coating of plastic or paper sheets. The oil booster, which resembles a steam ejector with an oil vapor boiler, generally filled with a light lube-type oil or a chlorinated hydrocarbon, doubles for the fourth stage of a steam ejector.

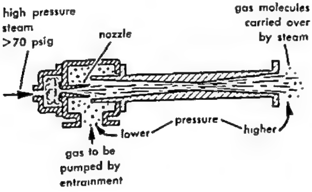


Fig. 3 Single-stage steam ejector pump

High vacuum pumps. The condensation or diffusion pump is used almost as widely as the mechanical backing pump. The simplest type is an inverted vapor stack (Fig. 4); refinements include means for continuous purification of the operating fluid Some of the many fluids are mercury, heavy petroleum oils, phthalic and sebacic esters, chlorinated hydrocarbons, and silicone oils. The latter have the unique property of not darkening or losing effectiveness when exposed hot, to the atmosphere. In all versions, a heater in the base vaporizes the filling, and the vapor streams out of the annular jets on the standpipe, entraining the entering gas before being condensed by the cooled walls of the pump. An efficient pump removes nearly half the entering molecules at first strike. Diffusion pumps suffer from back-streaming of the operating fluids

Fig 4. Diffusion high-vacuum pump

and are generally used with a trap cooled with solid carbon dioxide or liquid nitrogen to minimize this

The all-dry high-vacuum pump relies on the gettering or absorptive power of a reactive metal In one such pump, a titanium wire is vaporized by contact with white-hot tungsten The metallic vapor condenses on the cooled walls of the housing where each successive equivalent monolayer entraps residual gas Because the noble gases, particularly helium, are poorly pumped, a high voltage is sometimes applied to ionize them and confer a spurious affinity for titanium Depending on the gases being pumped, the ions combine with other materials to form a solid, become embedded in the container walls, or are attracted to charged plates away from the evacuated space These pumps maintain large-particle accelerators under hi acuum for pro-

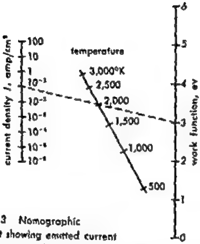


Fig 3 Nomographic chart showing emitted current density as a function of work function and temperature. This chart assumes unity coefficient in the emission equation. Current density indicated here must be multiplied by the actual emission coefficient to get true current density (From K R Spangenberg, *Fundamentals of Electron Devices*, McGraw-Hill, 1957)

gases) with the result that the vacuum becomes progressively better with time until it reaches a value of about one-billionth of atmospheric pressure At the same time that the getter is flashed, the emitter material of the cathode is heated by passing a current through the filament within it. This heat releases further gas, which is captured by the getter film At the same time the cathode is activated, and electron emission increases.

Emitters. Electron emission is required in all vacuum tubes Ordinarily, the cathode is heated by the passage of an electric current either through itself or through a filament in contact with it, and thermionic emission results Other types of emission are secondary emission, field emission, and photoemission

Various efficient emitters of electrons are available although these are surprisingly few These emitters must be chosen for different applications according to their characteristics The principal characteristics of emitters are (1) operating temperature (2) emission efficiency, (3) maximum emission capability, (4) resistance to back bombardment by positive ions, (5) insusceptibility to emission poisoning, (6) life (active and shelf), (7) resistance to mechanical shock, (8) ease of fabrication and (9) cost

Naturally, two of the most important characteristics are the total emission capability and the emission efficiency These characteristics are high if the emitter gives a large electron current at a relatively low temperature This in turn depends upon a factor known as the work function which is a measure of the height of the potential barrier that electrons within the emitter must scale in order to be liberated and freely emitted The work function is commonly given in units of electron volts and ranges from 1 to 5 electron volts.

Emitters fall into three classes: metals, atomic films, and compounds which are usually rare-earth oxides. Examples of each of these will be discussed.

Tungsten Of the pure metals, tungsten is one of the most extensively used as an emitter. This is not because it is such a good emitter, but rather because it is extremely rugged Tungsten has a work function of about 4 5 electron volts. It requires heating to a temperature of about 2200°K before it emits satisfactorily. Although tungsten can be heated to quite a high temperature (melting point is 3655°K), it is rarely operated above 2600°K because the evaporation at higher temperatures causes appreciable reduction in emitter life. The fact that tungsten can be operated at a higher temperature compensates to a large extent for the high work function, because the emission is equally sensitive to both.

Electron emission is given by the Richardson-Dushman equation which has the form

$$J = AT^2\epsilon^{-e\phi/kT}$$

where J = current density of emitted current, amp/cm²
A = constant of theoretical value, 120 amp/cm²
k = Boltzmann's constant, 1.38×10^{-23} watt sec
e = charge of the electron, 1.60×10^{-19} coulombs
ϕ = surface work function, electron volts (ev)
T = cathode temperature, °K

A nomograph of emitted current density as a function of temperature and work function is shown in Fig. 3

Cesium Cesium has one of the lowest work functions of all of the metals, on the order of 1.75 ev When cesium is deposited on other metals, the work function *may* be even lower. As a result cesium makes a good emitter However it is practical

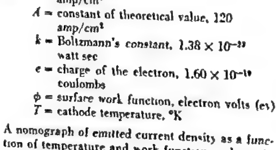


Fig 4 The emission of monatomic films on tungsten (After S. Dushman in K R Spangenberg, *Vacuum Tubes*, McGraw-Hill, 1948)

a control-grid electrode, in which case the tube is known as a triode, or three-element tube (*see* TRIODE, VACUUM). The interelectrode capacities between the control grid and the anode can be reduced by the insertion of a screen grid between the two, and the tube becomes a four-element, or tetrode, tube (*see* TETRODE, VACUUM). The anode current–anode voltage relationship is altered by the insertion of the screen grid, and problems in secondary electron emission from the anode (plate) arise and are suppressed by means of a suppressor grid *See* PENTODE, VACUUM.

Figure 1 is a sketch of a miniature pentode; triodes would appear similar except for the omission of the extra grids. Figure 2 is a collection of the standard symbols used for the various tube types. The beam power tetrode operates electrically just as if it possessed a suppressor grid and is occasionally referred to as a pentode.

Mechanical structure. The pentode tube of Fig 1 consists of an inner assembly of electrodes, made of nickel in receiving tubes, mounted on a base through which electrical connections are brought out, and enclosed in a glass or metal envelope which is evacuated The evacuation of the tube is necessary to obtain electron emission from the cathode, because emission does not occur at atmospheric pressure. The major portion of the manufacture is highly mechanized, but the assembly of the delicate inner workings is still done by hand.

Evacuation of the envelope. Vacuum tubes must be highly evacuated to secure a high degree of electron emission from the cathode over a long period of time. In manufacture, receiving tubes are evacuated with a mechanical pump Such pumps can easily achieve a pressure of 10^{-3} mm of mercury, or one-millionth of atmospheric pressure. This pressure is still too high by a factor of about 1000 for good operation, therefore tubes are further evacuated chemically by flashing a getter after they are sealed off as explained later. This deposits a chemically active substance on the inner walls of the tube, and this substance combines chemically with the major portion of the residual gases. The pressure in the tube before sealing is measured by a vacuum gage, usually of the thermocouple-type although other types such as the McLeod absolute-pressure gage, the Pirani gage, or the Philips gage can be used (*see* VACUUM MEASUREMENT).

Tubes must be completely fabricated before they can be evacuated. The electrodes are assembled on a base or stem structure to which a tube for exhausting the gases is attached. The glass envelope is placed over the electrode assembly like an inverted test tube. The envelope is then sealed to a glass skirt on the base stem by rotating the tube assembly and envelope in a ring of fires, which heat the envelope until it shrinks and seals to the base stem skirt. The excess glass from the lower portion of the envelope is then cut off.

Degassing of tubes. After the tube envelope is sealed to the base stem, the tube is evacuated by a mechanical pump. It is necessary to heat the tube and its electrodes to drive off gases that have gathered on the interior surfaces. This is done in part by heating the entire tube with a gas flame. The envelope must not be heated above 350°C by this means, or the glass will soften.

This is not a sufficiently high temperature to drive off gases that have collected on the electrode surfaces. Therefore, the electrodes are further heated by radio-frequency (rf) induction. This is done by putting a coil around the outside of the tube and passing a large rf current through the coil (*see* INDUCTION HEATING). The magnetic field produced by this current induces sufficiently large currents in the tube electrodes to heat them to a dull red heat It is then possible to evacuate the tube to one-millionth of atmospheric pressure, at which point the tube is sealed off and disconnected from the pump *See* VACUUM PUMP.

Getters. After the tube is partially evacuated and sealed off by the process described above, it is necessary to improve the vacuum, that is, decrease the pressure. This is done by flashing, or evaporating, a getter material onto the interior surface of the envelope. Barium is most extensively used as a getter material, although magnesium, calcium, sodium, and phosphorus have also been used. The getter material is usually contained in a rolled nickel tube, and is heated by radio-frequency induction so that the getter material vaporizes and escapes the crack in the tube. The vapor condenses as a metallic film on the inner wall of the envelope, giving rise to the shiny mirrorlike film clearly visible either at the base or at the tip of the tube When gas molecules come in contact with this layer they will combine (except for the noble

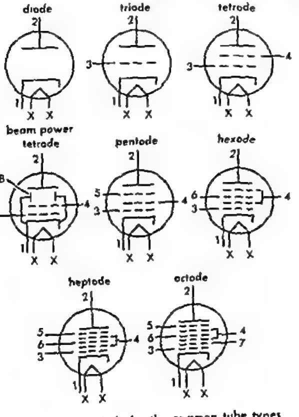


Fig 2. Circuit symbols for the common tube types

relationship between the anode voltage and anode current. For the triode the situation is complicated by the presence of the control grid, which introduces control-grid voltage as a third variable. In this case one of the three variables, anode voltage, anode current, or grid voltage, is held fixed at some value and a plot relating the other two variables is made. This plot is repeated for successive values for the fixed variable until a family of curves have been obtained If control-grid voltage is held fixed and the anode voltage and current are varied, the curves obtained are called plate characteristics (Fig 7) If anode voltage is held fixed and anode current plotted as a function of grid voltage, the curves are called grid plate transfer characteristics as shown in Fig 8 Curves obtained by holding anode current fixed and varying anode voltage as a function of grid voltage are known ...

the dc operating point for the tube when it is connected into a given circuit with given applied voltages (*see* AMPLIFIER) Once the dc operating point has been established for a tube operating as a small-signal amplifier, the only remaining need for the characteristic curves is that of determining the derivative or slope of each curve at the operating point These slopes are called the *ac parameters* of the tube, and they are used in amplifier-circuit design calculations

Amplification factor The amplification factor μ of a tube is one of its most important characteristics. This applies especially to triodes. The amplification factor is a measure of the voltage gain that could be achieved working into a very high load resistance. Actual voltage gain will normally be one-half or less of the value of the amplification factor.

The amplification factor of a tube is defined as the negative ratio of an increment of plate voltage e_b to an increment of grid voltage e_c required to keep the plate current i_b constant Thus if the grid is made 1 volt negative from its operating value and the plate must be made 30 volts positive from its operating value to restore the original current, then the amplification factor of the tube is 30 Expressed mathematically

$$\mu = -\left.\frac{\partial e_b}{\partial e_c}\right|_{i_b = \text{constant}}$$

Graphically, this equation states that μ is the negative slope of the constant current characteristics of Fig 9

The amplification factor of a tube is relatively independent of the operating voltages except in the vicinity of plate-current cutoff where it falls to a lower value. The amplification factor depends upon the geometry of the electrodes Factors that tend to give a larger amplification factor are larger grid wires, smaller grid-wire spacing and larger grid-plate separation Geometrically similar tubes of different sizes have the same amplification factor.

Typical values of μ run between 10-100 for modern triodes. The amplification factor is normally not used in calculations involving pentode tubes, therefore values for these tubes are seldom given by the manufacturers.

Mutual conductance. Mutual conductance, or transconductance, is defined as the ratio of the variation in plate current to the variation in grid voltage. Specifically, mutual conductance g_m is defined as the increment of plate current i_b divided by the increment of the control grid voltage e_c producing

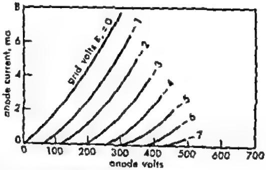


Fig 7 Typical plate characteristics for a triode vacuum tube

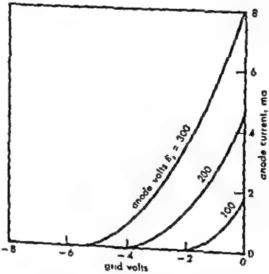


Fig 8 Typical grid-plate transfer characteristics

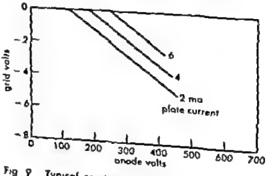


Fig 9 Typical constant-current characteristics

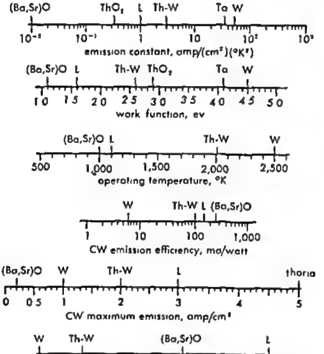


Fig 5. Relative characteristics of various emitters. *(From K. R Spangenberg, Fundamentals of Electron Devices, McGraw-Hill, 1957)*

only at low temperatures and is therefore restricted to use in phototubes. Unfortunately, it cannot be heated much without evaporating the cesium from the base layer and thus lowering the emission capability. Shown in the Fig 4 are the characteristics of various cesium (Cs) surfaces as a function of temperature.

Tantalum. Next to tungsten, tantalum is one of the most extensively used of the pure metals It has a slightly lower work function of about 4.1 ev Its melting temperature is about 100° lower than that of tungsten Tantalum is not as rugged as tungsten, but it has the advantage that it can be easily formed into sheets for making specially shaped cathodes.

Thoriated tungsten It is possible to obtain emission higher than that from pure metals from an atomic film of one metal on another. Of the various possible combinations, thorium on tungsten is the most extensively used. Such an arrangement has a work function of only 2 6 ev and gives rise to emission densities of about 2 amp/cm² compared with 1 amp/cm² for tungsten. The processing of thoriated tungsten filaments is involved, but it has been perfected. Usually such surfaces are operated at about 1800°K since the thorium coverage on the tungsten is reduced by *excessive evaporation* at higher temperatures. In building tubes using thoriated tungsten emitters it is necessary that the tubes be very highly evacuated, because a little residual gas can cause ionization which will gradually destroy the emission.

Oxide coatings. The most extensively used of all the emitters are the oxide emitters, which are usually a mixture of barium and strontium oxides. Discovered at the beginning of the century, they have been under development and study ever since.

Oxide coatings commonly have a work function of the order of 1.1 ev. They can be operated at temperatures of 1050°K. They yield current densities on the order of 0.2 amp/cm² for thousands of hours or densities on the order of amperes per square cm for shorter periods of time. This combination of characteristics puts the oxide emitter in great demand. This is not all without some limitations, however, because the oxide emitter is probably the least rugged of all of the emitters. It is most easily poisoned by traces of gas and is most difficult to activate in the first place.

Relative capabilities of various emitters Figure 5 is a set of scales showing the relative characteristics of the various emitters. This set of scales shows that no one emitter is superior in all of the characteristics that are of usual interest.

Another comparison of the capabilities of emitters is shown in Fig. 6. This figure shows the emission current density as a function of the heating power of typical filaments in watts Also shown in this are contours of constant efficiency in milliamperes per watt. In these figures the L, or Lemmen's cathode, is essentially a barium oxide-tungsten emitter made by compressing tungsten powder and barium oxide into a hard mass which can be handled in air and machined.

Tube coefficients. The operation of the diode tube can be described completely by a graph of the

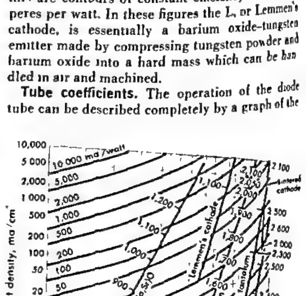


Fig. 6 Relative emission efficiencies of various cathodes *(From K. R Spangenberg, Fundamentals of Electron Devices, McGraw-Hill, 1957*

e dissipated on the plate or the maximum extent to which the plate electrode can be heated. This results from the fact that for different types of service there is a maximum efficiency that can be obtained, such as about 25% for amplification of speech, which means that three-fourths of the power into the tube (exclusive of filament power) is dissipated at the plate and the remaining one-fourth is transmitted as useful output. Higher efficiencies such as 50% for class B, or half sine-wave operation, and 65% for class C, or radio-frequency amplification, can be obtained.

The ability to dissipate power at the plate will depend upon the construction of the tube and the method of cooling, in particular. For ordinary receiving tubes, with no special means of cooling, the plate dissipation may run a few tenths of a watt per square centimeter. If the tube is built so that forced-air cooling may be used on an external anode, the dissipation may be increased to about 1 watt/cm². Low- and medium-powered transmitting tubes often have anodes or plates that run red hot and are cooled by radiation. Such plates may have a dissipation capability of 4–10 watts/cm². Larger tubes will often be water cooled over an external plate and will have dissipation capacities of 30–110 watts/cm².

Interelectrode capacitance. All tubes have an electrical capacitance between their electrodes which influences the tube behavior in amplifier and oscillator circuits. The simplest case is the triode. Here there are three active elements, the cathode, the control grid, and the plate. There are three interelectrode capacitances, one formed between each electrode and the other two, with the capacitance from electrode one to two being the same as that from two to one. The cathode to control-grid capacitance is the largest and can be estimated from the usual formulas for capacitors by inserting the area of the control grid. The capacitance from control grid to plate is next largest and can be calculated in the same fashion. There is also a capacitance from cathode to plate even though the control grid exists between them. This capacitance corresponds to the electric flux lines which thread the control grid in going from cathode to plate *without terminating on the control grid*. This capacitance is the smallest of the three, but is still large enough to interfere with amplifier operation. The capacitance measured from control grid to ground forms part of the input capacitance and that measured from plate to ground forms the output capacitance. These are most important in determining the product of the voltage gain A and the bandwidth B of the tube. This product is proportional to the ratio of the mutual conductance g_m to the sum of the input and output capacitances C_o.

$$AB = \frac{g_m}{2\pi C_o}$$

In order to secure a large gain-bandwidth product it is naturally desirable to have these input and output capacitances as small as possible.

In addition to the purely electrostatic capacitances that exist in the tube, a component of input capacitance is generated by the plate-circuit action ...

... grid-to-plate capacitance multiplied by the voltage gain plus 1. The complete input capacitance becomes

$$C_{in} = C_{gk} + C_{gs} + C_{gp}(1 + A \cos \theta) \quad \text{micromicrofarads } (\mu\mu f)$$

where C_{gk} is the control-grid to cath... tance, C_{gs} is ... tance, C_{gp} is ... A is the magnitude of the voltage gain of the stage, and θ is the phase angle. In the case of a triode, C_{gs} is zero.

In view of this relationship it is obviously desirable that the grid-to-plate capacitance be as low as possible. For this reason pentodes are preferred to triodes, because in the pentode there is shielding between the grid and plate with the result that the grid to plate capacitance is ordinarily less than one-tenth of that encountered in triodes.

The relative values of the different interelectrode capacitances can be indicated by a few examples. In the 6AB4 high-mu triode the input and output capacitances are 2.2 and 0.5 $\mu\mu f$, respectively. The grid to plate capacitance is 1.5 $\mu\mu f$. In the 6AC7

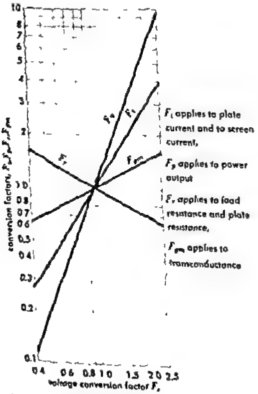


Fig. 10. Voltage conversion factors. (From K. R. Spangenberg, *Vacuum Tubes*, McGraw-Hill, 1948)

it for a constant value of plate voltage e_b. Mathematically

$$g_m = \left.\frac{\partial i_b}{\partial e_c}\right|_{e_b = \text{constant}}$$

This is the slope of the curves of the transfer characteristics of Fig. 8. This ratio has the dimensions of conductance, and the adjective "mutual" or prefix "trans" is applied because it relates variations in the grid circuit to changes in the plate circuit.

In tubes other than triodes the term transconductance rather than mutual conductance is used, and it is necessary to specify the two electrodes involved since more than one combination is possible. However the same symbol g_m and definition apply. Modern pentodes have transconductance between 800 and 5000 micromhos and going up to about 10,000 for special high-gain pentodes.

Mutual conductance and transconductance are among the so-called constants of a tube. Actually these conductances are constant only for a given set of operating voltages. They do vary with the operating voltages and currents. In general, the mutual or control-grid transconductance will vary as the cube root of the plate current; if the plate current is increased by a factor of 8, the mutual or transconductance will go up by a factor of 2. The mutual or transconductance will also depend upon the tube geometry, particularly upon the dimensions of the control grid. The mutual conductance also varies inversely as the 4/3 power of the cathode-grid spacing.

Dynamic plate resistance. The dynamic plate resistance is a third characteristic commonly cited to give a measure of tube performance. It is essentially the resistance offered to the flow of an alternating component of current through the tube. Specifically, it is defined as the ratio of an increment of plate voltage e_b to an increment of plate current i_b for a constant grid voltage e_c. Expressed mathematically,

$$r_p = \left.\frac{\partial e_b}{\partial i_b}\right|_{e_c = \text{constant}}$$

The dynamic plate resistance is the slope of the curves of the plate characteristics of Fig. 7. The dynamic plate resistance of triodes is commonly in the range of 1000–20,000 ohms, while that of pentodes is commonly of the order of several megohms.

The dynamic plate resistance tends to vary inversely as the cube root of plate current. The dynamic plate resistance of a tube is a measure of the output impedance of the tube. As mentioned above, it is relatively low for triodes and high for pentodes.

Relation between tube coefficients. If the definitions of the three so-called tube constants previously discussed are examined carefully, it will be seen that there is a simple relation between them. Specifically, the amplification factor μ is equal to the product of the mutual conductance g_m and the dynamic plate resistance r_p; that is,

$$\mu = g_m r_p$$

Since the amplification factor of a tube is relatively constant, this means that the dynamic plate resistance of a tube decreases as the mutual conductance increases, and vice versa.

Variation of characteristics. Vacuum tube characteristics are subject to considerable change with operating voltage. This is explained principally by Child's law, which states that the current drawn from a cathode varies approximately as the 3/2 power of the equivalent voltage. As a consequence of this, the mutual conductance varies as the 1/3 power of the current and as the square root of the voltage. Thus if a tube is operated with its anode voltage about 40% above normal, the mutual conductance will be 10% higher than normal. In a similar fashion, the plate resistance will vary inversely as the square root of the voltage. The electrode currents themselves will vary as the 3/2 power of the electrode voltage. The power output and electrode dissipation will vary as the 5/2 power of the electrode voltages. These relations are indicated in Fig. 10. The above assumes that the voltage variations are not large enough to impair the emission. The relations given above apply only when the cathode emission is sufficient so that the current is space-charge limited rather than emission- or temperature-limited. *See* LANGMUIR-CHILD LAW; SPACE CHARGE.

However, the emission characteristics are influenced by filament voltage variation. The principal factors that are affected are the emission, the temperature, and the life. The peak emission capability of a cathode is found to vary as the 7.5 power of the filament voltage. This is a very rapid or sensitive function of the filament voltage. The temperature of the cathode itself varies only as about the one-tenth power of the filament voltage. On the other hand, the life of a tube varies inversely as the fifteenth power of the filament voltage. This means that if a tube is operated with its filament voltage 15% higher than rated, the life can be expected to be only one-tenth of the rated value. Conversely, the life of a tube can be doubled if its filament is run at 91% of the rated value.

Tube selection. The circuit designer is confronted with a bewildering array of available tubes. There are several hundred tube types in common use and a total listing of available tubes would probably run over 1000. Fortunately, the principal tube manufacturers publish tube manuals, which list the characteristics of the principal tubes and recommend tubes for different types of service. Even so, the circuit designer will ordinarily require considerable experience before he can choose the best tube for his purposes from the large number available.

Plate dissipation. The power output of a vacuum tube will generally be limited by the plate dissipation, that is, the maxi[illegible] ower that can

begins to show up at about 50 Mc and can be represented by an additional resistance between the grid and ground, or cathode, of the tube. The resistance varies inversely with frequency and typically has a value of about 1000 ohms at 100 Mc.

Reliability. In general, the term reliable needs to be defined for each application and involves a combination of length of life, mechanical ruggedness, and uniformity of characteristics from tube to tube. More specifically, a reliable tube is one which has a high probability that it will operate normally when taken from stock and installed in equipment for which it was intended and a subsequent low probability that it will fail during subsequent operation in that equipment for some fixed period of time.

Causes of failure. There are many causes of failure in vacuum tubes. Ordinary tube life is considered to be 1000-3000 hours. Long-life tubes are generally expected to operate 50,000 hours at least.

Heater failure is a primary cause of tube failure. Filaments burn out because of mechanical strains resulting from turning the filament on and off. There may also be heater-to-cathode shorts because of failure of insulation.

A second cause of failure is conductor breaks. These include the opening of welds at joints between conductors and nonweld breaks which are something of a mystery although they are as common as the breaks at the welded joints.

A third cause of failure is gas leakage. The vacuum becomes poor because gas or air has leaked into the tube. It may result from electrolysis of glass at the leads or all kinds of strains.

A fourth type of failure is grid emission. This occurs after a longer period of time and results from emitting material being transferred from the cathode to the grid which then begins to emit electrons, spoiling the normal operating characteristics of the tube.

A fifth cause of failure is emission failure. Usually this will be significant only in tubes running more than 1000 hours. Emission failure may result from gas poisoning by traces of chlorine, sulfur, oxygen, water vapor, or carbon dioxide within the tube. It may also result from exhaustion of the emitting material caused mainly by a combination of the barium with reducing elements in the base nickel. This tendency can be reduced by using smaller amounts of reducing material, such as silicon, titanium, and magnesium, in the base nickel. For long-life tubes very pure electrolytic nickel is used as a base on which the oxide emitting material is deposited. Also, an interface layer of barium silicate and similar compounds may form at the metal-oxide junction. These compounds have a higher resistance than the oxide and reduce the emission. An interesting variation of this is the tendency for this interface layer to build up faster when current is not drawn than when the tubes are in operation.

Failure statistics. As a general rule the number of tubes surviving in a given lot follows what is nearly a negative exponential law after the first 50–500 hours. Therefore, a curve can be drawn and projected to estimate the number of tubes that will remain after a given period of time.

It is more difficult to estimate the life of a particular tube because so many causes of failure are involved that the general statistics do not apply. If, however, the tube survives mechanical failures and the like and its life is dependent upon emission, then the remaining life can be estimated by observing the emission under conditions of reduced temperature of the filament. The lower the temperature at which ...

... failures are initially high but drop off with time. These will ordinarily account for 95% of the fail...
the
first
increases linearly with time and is significant only after the first 1000 hours. Ultimately, emission failures will take over, causing the over-all failure rate to rise again.

Due to this combination of causes from different sources, the rate of failure ...
sembles
short pe...
mechanical failures are high but drop off as the tube grows older. Then there is a period of middle life, during which the failure rate is more or less uniform. As age advances failures due to leakage and emission cause the mortality rate to increase again, corresponding to the problems of old age in the human mortality picture.

As previously mentioned, the survival of tubes after the initial period tends to be nearly a negative exponential curve. For this to be entirely true a constant percentage of the tubes must fail per unit of time. This is approximately true except for the initial run-in period.

The individual performance of tubes can be predicted only to the extent that the emission is involved. In order to do even this it is necessary to operate the tube with a reduced filament voltage and observe the effect upon the emitted current. Such procedures have been formalized into what is known as marginal checking. A tube with considerable emission life remaining can be subjected to a greater reduction of filament voltage without an appreciable decrease in emission current then can a tube with a shorter remaining emission life.

Tube sockets. Almost all vacuum tubes are built with a number of pins projecting from the base. These pins are plugged into a socket, allowing easy replacement of tubes. Small tubes have simply straight wires projecting from the base of the tube. Larger tubes have small cylinders ...
which ...
been ...
tial p...
larger ... type tubes require provision for the high currents involved; these often have

sharp-cutoff pentode the input capacitance is 11 $\mu\mu$f, the output capacitance is 5 $\mu\mu$f, and the number-one grid to plate capacitance is 0.015 $\mu\mu$f. Thus it is seen that in the pentode the input and output capacitances are higher than for the triode, but the interelectrode capacitance is much lower. The former factor contributes to a slight reduction in the gain-bandwidth capabilities, but the latter offers the possibility of much higher gain without oscillation or instability.

Frequency limitations. All electron devices, including tubes, are subject to various frequency limitations. Some limitations are related to the circuits, and some are related to the movement of the charge carriers through the device. In general, the higher the frequency the harder it is to obtain appreciable power. Special approaches may be used to cover extremely high frequencies, but these generally result in narrow-band devices that do not cover the lower frequencies.

Circuit reactance. One of the inherent limitations in vacuum-tube amplifiers is the associated circuit reactance. As amplifiers are built to achieve a wide band of operation, it is generally found that there is a high-frequency limit related to a resonance between circuit inductance and circuit capacitance. These can be reduced only so much until the capacitance is limited by the input and output capacitance of the tubes. A general consequence of this is that there is a more or less constant product of voltage gain and bandwidth which is characteristic of each tube. If an effort is made to increase the voltage gain by changing circuit values, it is found that the bandwidth is decreased, and vice versa. The best pentode-type tubes will have gain-bandwidth products of the order of 100 or 200 megacycles.

Tube reactance. Several tube reactance effects are conspicuous in amplifiers operating at high frequencies. The first of these is the so-called Miller effect, previously discussed, which results from a feed-back of the output voltage to the input of the tube through the grid-to-plate interelectrode capacitance and has the effect of increasing the input capacitance of the tube. Another effect, which is found in tubes and which affects the high-frequency behavior, is associated with the cathode lead inductance of the tube. This introduces inductive impedance into the input circuit. Specifically, the cathode lead inductance causes the existence of an input component of conductance that is proportional to the square of frequency and also proportional to the cathode lead inductance, the cathode grid capacitance, and the mutual conductance of the tube. Since this input conductance grows as the square of frequency, it is a serious factor in limiting high-frequency operation. A similar effect, caused by the transit time of electrons within the tube, also introduces a component of input conductance proportional to the square of frequency and to the mutual conductance of the tube.

Electron transit time. As frequency is increased, the transit time of the electrons within the tube is a progressively larger fraction of the radio-frequency cycle. This has some important consequences, which actually affect the external impedance of the tube as measured at the input terminals. Specifically, the transit time gives rise to an input component of conductance that varies as the square of the frequency, the mutual conductance, and the square of the total transit time. This is the same kind of frequency variation as that caused by cathode lead inductance with the result that the two effects are often hard to separate. Typical values of input resistance due to this effect are of the order of 1000 ohms at 100 megacycles (Mc).

Noise. Noise is the ultimate limitation of almost any kind of amplifying device. Theoretically, and also practically, it is possible to make amplifiers which have gains as high as desired. However as the amplifier will always have a certain amount of internally generated noise, which will mask signals below a certain level, there is a limit to the smallest signal that may be amplified. The noise in vacuum tubes is due to the emission of electrons from the cathode in a random fashion. Apparently, the electrons come off in bursts or in a varying amount rather than continuously. This gives rise to a noise, which though small is still large enough to be a limiting factor.

Noise is also found in resistors due to the random motion of electrons resulting from molecular agitation. Noise evidences itself as a voltage e across the resistor. This voltage is proportional to the square root of the absolute temperature T, the resistance value R, and to the width of the frequency band B over which the observation is made

$$e = \sqrt{4kTRB} \quad \text{rms volts}$$

where k is Boltzmann's constant (1.38×10^{-23} watt-second/°K).

In considering tube noise it is convenient to specify this in terms of an equivalent noisy resistor which, if put in the grid circuit, would develop as much noise as does the tube in question. Although the phenomena involved are quite complex, the resulting relationship is quite simple. Specifically, for triodes, the equivalent noisy resistor in ohms for the grid circuit is simply equal to 2.5 divided by the mutual conductance in mhos of the tube. For a tube with a mutual conductance of 10,000 micromhos, this would indicate an equivalent noisy resistance of 250 ohms. This relationship is observed almost perfectly for practical triodes with equivalent resistances covering the range of 200–2000 ohms.

The triode is the quietest of all vacuum tubes. Pentodes are noisier by a factor of approximately 4, which means that as a rough guide the equivalent noisy resistance is equal to 10 divided by the mutual conductance of the tube. This additional noise arises from a random division of electrons between the screen and the plate of the tube.

At high frequencies, additional components of noise need to be considered. T[illegible]ional noise

in reflected current more than offsets the increase in directly intercepted space current that is taken on by the screen grid as a result of its more positive potential.

The two dotted curves of Fig 12 are for differences of the number 2 and number 3 potential of 54 and 90 volts, respectively The magnitude of the negative resistance made available by this means is on the order of 3500 ohms This is considerably less than that obtainable from a dynatron, which is usually on the order of 10,000 ohms The region of negative resistance is limited at low voltages by the condition that the suppressor grid is returning all the electrons that approach it Beyond this condition the suppressor grid has virtually no influence Correspondingly, the region of negative resistance is limited at high voltages by the condition that the suppressor grid is passing all the electrons that approach it and so again loses control

In actual applications, the screen and suppressor grids are separately biased and fed through separate resistors but are coupled by a large capacitance connected directly across the tube leads This means that the number 2 and 3 grids are connected together as far as voltage variations are concerned over a large band of frequencies The negative-resistance characteristic is available from low audio frequencies, dependent upon the size of the coupling capacitor compared with the size of the resistors in series with the electrodes, to frequencies on the order of 60 Mc, at which transit-time effects disturb the relations With proper connections the negative screen resistance of a pentode can be made to furnish either sinusoidal or square waves Likewise trigger and flip-flop characteristics can be made available

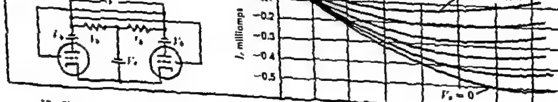
Push pull negative-resistance circuit It is possible to connect two triodes or two pentodes in a push pull arrangement to obtain an excellent negative-resistance characteristic The circuit and resultant characteristics are shown in Fig 12 The current that flows through the input terminals shown consists of two components: (1) that produced by the applied voltage, which is in one direction, and (2) that produced by the tubes, which will be in the opposite direction because of the cross connection of the grids The latter component of current can be made much larger than the former by tapping sufficiently high on the plate resistor of the tube, usually across the entire resistor. A negative-resistance characteristic, as shown, results. The resistance available has the approximate value of

$$R = \frac{2r_p}{(r_b + r_p)/r_b - \mu k}$$

For small values of impressed voltage, where R is the effective value of the resistance at the input terminals, r_p is the dynamic plate resistance of the tubes, r_b is the value of the plate resistor, μ is the amplification factor of the tubes, and k is the fraction of the voltage developed across the plate resistors that is applied to the other tube If k is sufficiently large, the effective resistance will usually be negative

In application, the cross connection between grids is made through a large capacitance If k is made too small, the circuit will neither oscillate nor produce square waves. As k is raised, the circuit will first produce sinusoidal waves and then square waves as k is made larger The negative-resistance characteristic is available from low audio frequencies to frequencies on the order of megacycles. This is the basic circuit of the Eccles-Jordan trigger circuit and multivibrator. *See* MULTIVIBRATOR

Electron-ray indicator tube This tube is capable of visual indication of potentials between 0 and about 100 volts. The indication is in the form of an occluded sector of an otherwise fluorescent, circular target. The arc of the occluded sector is nearly linearly proportional to control electrode voltage with a maximum arc length of about 100 degrees when the control voltage is zero Figure 13 depicts



12 Charac... negative-resistance circuit. (From K R Spangenberg, Vac...

bars or blades fitting into slots of a socket, resembling a knife-blade construction.

Tube connections are obviously important to ensure proper operation. Any tube manual shows the connections for the various tubes. These are invariably base or socket connections as seen from below

Tube sockets are fabricated of various plastics and ceramics selected for their low electrical conductivity at high frequencies, their ability to withstand high operating temperatures, and for their mechanical strength. Electric shields are often incorporated into tube sockets. A shield is often provided between the contacts at the bottom of the socket, as well as partial or complete tube shields above the socket Since reduced bulb temperature contributes to longer tube life, many shields are designed to transfer tube heat away from the tube by conduction and radiation.

Special tubes. Special tubes fall into two groups: (1) those which are designed for some unique purpose and (2) those which are standard tubes of assorted types that have been enclosed within a common envelope The latter group are produced to the express requirements of large manufacturers, who find it possible to cut costs by using envelopes which include several individual tubes Typical names for such tubes are duodiode (two diodes), duodiode–high-mu triode (two diodes plus a triode having a high amplification factor), twin triode (two triodes), and so forth.

The first group includes many common types of tubes that are considered special for one reason or another, such as a screen grid that is capable of unusually high power dissipation, or a mechanical structure that permits a spring-mounted anode to move under tube acceleration and thus alter the characteristics of the tube (accelerometer tube), as well as tubes that are designed for one unique purpose Tubes of this type would include the electron-ray indicator tube and negative-resistance tubes, which are covered here, as well as television camera and projection tubes, cathode-ray oscilloscope tubes, and ultra-high-frequency tubes *See* BACKWARD-WAVE TUBE, CATHODE-RAY TUBE, KINESCOPE, KLYSTRON, MICROWAVE TUBE, PHOTOTUBE, STORAGE TUBE; TELEVISION CAMERA TUBE, TRAVELING-WAVE TUBE; X-RAY TUBE.

Dynatrons The negative resistance that is available over part of the plate current–plate voltage characteristic of an ordinary screen-grid tube is known as a dynatron characteristic. The negative-resistance characteristic results from the transfer of secondary electrons from plate to screen. When the screen grid is more positive than the plate, an increase in plate voltage will attract more primary electrons to the plate. However, relatively more secondary electrons are lost to the screen grid, so that the net plate current decreases rather than increases.

If a parallel resonant circuit is placed in the plate circuit of a screen-grid tube and the tube operated at voltages that give a negative-resistance characteristic, oscillations will occur in the plate circuit provided that the magnitude of the resistance of the parallel resonant circuit is greater than the magnitude of the negative resistance of the tube plate circuit. Oscillations will, in general build up to the point where the magnitude of the negative resistance as averaged over the cycle of oscillation equals the positive resistance of the parallel resonant circuit

The negative-resistance characteristic obtainable from a screen-grid tube is subject to change as the tube ages and as the secondary-emission characteristics of the plate change from any of a number of causes. For this reason, this type of negative resistance is not extensively used

Negative-resistance screen-grid pentode The screen grid of an ordinary pentode exhibits a negative-resistance characteristic if it is connected to the suppressor grid in such a way that an increase in screen voltage is accompanied by an equal increase in suppressor grid voltage This is evident from the curves of Fig 11 This family of curves shows the I_2–V_2 characteristics of a pentode for various values of V_3, where the numerical subscripts refer to the grid number in order from the cathode to plate. The solid curves show the I_2–V_2 characteristics As the number 3 (suppressor) grid is made more negative, the number 2 (screen) grid current decreases. If the number 2 and 3 grids are connected so that there is a constant difference of potential between them, the dotted curves shown in Fig 11 result. The screen current decreases as the suppressor grid is made more positive, because the latter then transmits a greater fraction of the space current that approaches it and, as a result, less current is returned to the screen grid This decrease

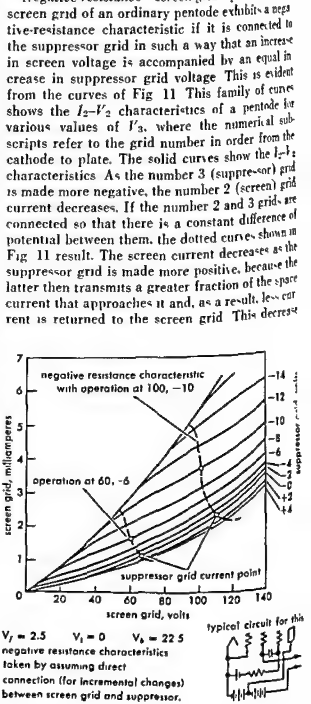


Fig. 11. Screen grid current-voltage characteristics of a pentode

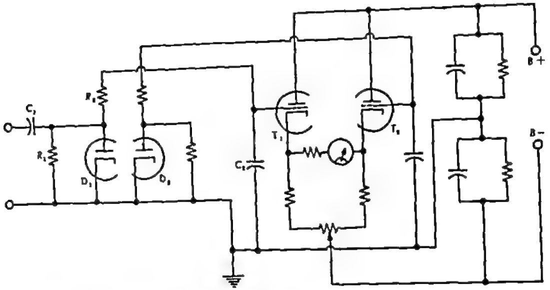


Fig 1 Circuit diagram of a diode rectifier–amplifier meter

warmup and heater-voltage variations Plate-voltage fluctuations are also balanced out by the amplifier, so that stable operation down to 1.5 volts full scale can be obtained without employing a regulated power supply The cathode-circuit resistance is high enough to stabilize the gain of the amplifier The constants of the rectifier circuit are chosen so that the diode operates on the exponential cutoff part of its characteristic This varies only slightly from tube to tube so that a new calibration is not required when the diode is replaced

High frequency performance For diode rectifier-amplifier meters this factor depends on the geometry of the diode and of the probe in which it is mounted The plate-to-cathode spacing determines the electron transit-time error, which is important at high frequencies and low voltages The series resonant frequency of the input loop of the rectifier determines the resonance error, which is independent of voltage The two effects combine to give a high-frequency error depending both on the frequency and on the scale range of the instrument. A set of error curves for a commercial instrument is given in Fig 2

Input impedance. At low frequencies the input resistance of a diode-rectifier meter is determined by the diode current and is effectively about one-fifth the value of the discharge resistance (R_1 in Fig 1), or of the order of 10 megohms Dielectric losses are a controlling factor at high frequencies and give a value of roughly 50–100 kΩ at 100 Mc in high-quality instruments This is in parallel with a capacitance of a few micromicrofarads

Direct-current vacuum-tube voltmeters have an input resistance determined by the grid current of the input tube. For ordinary receiving tubes this is

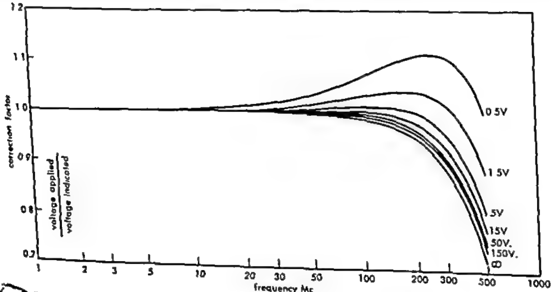


Fig. 2. Error curves of a commercial diode rectifier–amplifier voltmeter (General Radio Co)

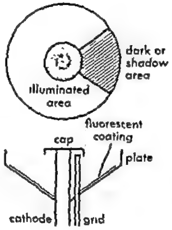


Fig. 13 The electron-ray tube (From K. R. Spangenberg, Vacuum Tubes, McGraw-Hill, 1948)

the typical construction of such an indicator tube.

Electrons emitted by the cathode are accelerated to the anode, or target, where they give up their kinetic energy, and cause a coating on the anode to fluoresce At one side of the cathode a single grid or ray-control electrode is located. If electrons are to pass by this electrode essentially undeflected, the electrode must be maintained at some potential between zero and plate potential For any voltage lower than this potential, the electrons in the vicinity of the electrode will be deflected away from it, thus giving rise to an occluded sector or shadow on the target.

The 6AF6-G is an example of this type of electron-ray indicator tube. More commonly, the indicator tube is found combined with a standard triode voltage amplifier tube in the same envelope The plate of the triode is usually directly coupled to the ray-control electrode The purpose of the triode is to furnish the necessary voltage gain to operate the ray-control electrode from lower level signals. Such tubes are sometimes referred to as magic-eye tubes [K R S]

Bibliography: W G. Dow, *Fundamentals of Engineering Electronics*, 2d ed, 1952, K R. Spangenberg, *Fundamentals of Electron Devices*, 1957; K R Spangenberg, *Vacuum Tubes*, 1948; J Yarwood, *High Vacuum Technique*, 3d ed, 1955

Vacuum-tube voltmeter

Any of several types of instrument in which vacuum tubes, acting as amplifiers or rectifiers, are used in circuits for the measurement of ac or dc voltage. The various types of ac vacuum-tube voltmeter all derive from the voltage to be measured a direct current to operate an indicating meter The vacuum-tube voltmeter is the principal instrument for voltage measurements at high frequencies, as the frequency limit is determined essentially by the rectifier characteristics. It is also the usual test instrument in a wide variety of applications where little power can be taken from the source. For other types of voltmeters, *see* VOLTMETER

Types. Six principal types of vacuum tube voltmeter are described briefly here. The most important type, the diode rectifier-amplifier meter, is discussed in detail in a later section

Plate-circuit rectification meter. In this meter the signal voltage is applied to the control grid of a tube, and rectification takes place because of the form of the grid voltage–plate current characteristic. The increase in average plate current operates the indicating meter. This was the first type used but is now rarely seen because the calibration depends greatly on the tube characteristics and the frequency range is limited in comparison with diode voltmeters A modification, called a reflex vacuum-tube voltmeter, uses the rectified voltage in the plate circuit to increase the negative grid bias. This permits much larger voltages to be handled without the grid drawing current and, at the higher voltage ranges, makes the calibration more linear and less dependent on tube characteristics

Grid-rectification meter. The grid and cathode of a tube act as a diode rectifier, and the rectified grid voltage, amplified by the tube, operates a meter in the plate circuit

Diode rectifier–amplifier meter. This is probably the most widely used meter. It has separate tubes for the rectification and dc amplification functions, permitting an optimum design for each The dc amplifier, which may have more than one tube, usually employs inverse feedback for the stabilization of gain (*see* DIRECT-COUPLED AMPLIFIER). The diode rectifier can be specially designed for high frequency performance and is usually mounted in a small probe which can be put at the point in the circuit where the voltage is to be measured

Dc vacuum-tube voltmeter. This is essentially the amplifying and indicating portions of the diode rectifier–amplifier meter, which are usually designed so that the diode rectifier can be disconnected for dc measurements.

Amplifier-type meter This meter has an ac amplifier, with gain stabilized by inverse feedback, preceding the rectifier.

Slide-back voltmeter. This meter employs a sharp-cutoff tube with adjustable negative grid bias The bias at which plate current just commences is observed with and without the unknown voltage, and the difference in bias, read on a dc voltmeter, is taken as a measure of the applied ac voltage.

Diode rectifier–amplifier meter. A simplified circuit diagram of a typical diode rectifier–amplifier instrument is given in Fig 1 The components to the left comprise a parallel-type diode rectifier circuit, in which the diode D_1 charges the capacitor C_1 to a voltage approximately equal to the peak value of the applied ac voltage, the rectified voltage appearing also across the diode D_1 The resistor R_1 permits C_1 to discharge when the voltage is removed To avoid low frequency error the product ωR_1C_1 should be about 100 or greater at the lowest frequency to be measured. The resistor R_2 and the capacitor C_2 filter out the ac component and the dc is applied to tube T_1 of a balanced two-tube dc amplifier A second diode D_2 in the same envelope with D_1 is connected to the grid of the second amplifier tube T_2 t alance zero drift from

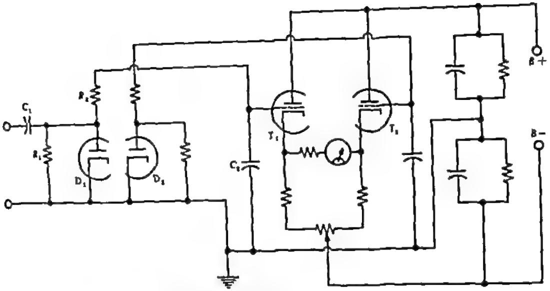


Fig 1. Circuit diagram of a diode rectifier–amplifier meter

warmup and heater-voltage variations. Plate-voltage fluctuations are also balanced out by the amplifier, so that stable operation down to 1.5 volts full scale can be obtained without employing a regulated power supply. The cathode-circuit resistance is high enough to stabilize the gain of the amplifier. The constants of the rectifier circuit are chosen so that the diode operates on the exponential-cutoff part of its characteristic. This varies only slightly from tube to tube, so that a new calibration is not required when the diode is replaced.

High frequency performance. For diode rectifier-amplifier meters this factor depends on the geometry of the diode and of the probe in which it is mounted. The plate-to-cathode spacing determines the electron transit-time error, which is important at high frequencies and low voltages. The series-resonant frequency of the input loop of the rectifier determines the resonance error, which is independent of voltage. The two effects combine to give a high-frequency error depending both on the frequency and on the scale range of the instrument. A set of error curves for a commercial instrument is given in Fig. 2.

Input impedance. At low frequencies the input resistance of a diode-rectifier meter is determined by the diode current and is effectively about one-fifth the value of the discharge resistance (R_1 in Fig. 1), or of the order of 10 megohms. Dielectric losses are a controlling factor at high frequencies and give a value of roughly 50–100 kΩ at 100 Mc in high-quality instruments. This is in parallel with a capacitance of a few micromicrofarads.

Direct-current vacuum-tube voltmeters have an input resistance determined by the grid current of the input tube. For ordinary receiving tubes this is

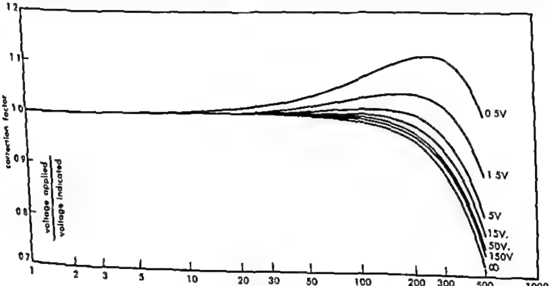


Fig 2. Error curves of a commercial diode rectifier–amplifier voltmeter (General Radio Co.)

in the range 10^{-7} to 10^{-9} amp, and the usual dc vacuum-tube voltmeter may have an input resistance of 10–100 megohms. Direct-current vacuum-tube voltmeters with special low-grid-current input tubes are called electrometer voltmeters, or vacuum-tube electrometers, and may have grid currents down to 10^{-15} amp or lower. Such voltmeters may be used with input resistances as high as 10^{11} or 10^{12} ohms.

Low-voltage limits. Diode rectifier–amplifier meters are limited in sensitivity by the square-law response of the rectifier at low voltages and by zero instability of the dc amplifier from contact potential variations in the diode and the first amplifier.

The maximum practicable sensitivity of a dc vacuum-tube amplifier with careful design is about 30 mv full scale, and electrometer voltmeters of this sensitivity are commercially available. A rectified voltage of this value requires an ac input of about 0.1 volt, and the smallest detectable ac voltage would be 5–10 mv. However, this limit is hardly feasible in practice, and full-scale sensitivities lower than about 0.5 volt on either ac or dc ranges are seldom seen in general-purpose instruments

In amplifier-type ac meters, however, the amplifier is ahead of the rectifier and measurements can be made down to about 1 mv over a 5-Mc bandwidth with full accuracy. Highly selective amplifiers, not usually classed as vacuum-tube voltmeters, are necessary to make ac measurements down to the microvolt level.

Waveform error. At voltages above about 5 volts, diode rectifier voltmeters respond essentially to the peak value of the applied voltage. The calibration assumes sinusoidal waveform. On other waveforms the meter reads 0.707 times the peak value, which may deviate by as much as the percentage of harmonics present from the rms value or from the value of the fundamental component Below about 0.3 volt the response of diode rectifier meters is essentially square-law, and the waveform error is negligible. Amplifier-type voltmeters provide ample power to operate the rectifier, so that it can be designed for linear response, giving reduced waveform error. *See* CURRENT MEASUREMENT, DIGITAL VOLTMETER; VOLTAGE MEASUREMENT. [W N T]

Vagina

In human beings, the fibromuscular tube extending from the external female genitals to attach around the cervix, the lower narrowed portion of the uterus. When relaxed, the vagina is 3½–4 in long and the walls are collapsed. It passes upward and backward to join the forward-tilted uterus at a right angle. The anterior wall lies below the bladder and urethra; the posterior wall, in front of the rectum. The mucous membrane lining contains two Bartholin's glands near the orifice which secrete a lubricant fluid. Although ordinarily folded and collapsed, the vagina has a great capacity for dilation during delivery, when it serves as the lower portion of the birth canal *See* UTERUS.

Various forms of the vagina exist in most mammals, but in lower forms the equivalent is either the cloaca, a common opening for urogenital and gastrointestinal systems, or specialized genital pores. [E.C ST]

Vaginal disorders

Congenital defects, inflammations, tumors, trauma, and functional disorders of the vagina, the musculomembranous canal from the vulvar opening to the uterine cervix. *See* VAGINA.

Congenital defects are uncommon but may include absence of the vagina, failure of its growth, or abnormal formations, such as double vagina. Congenital cysts, although common, are usually unimportant.

Inflammations of the vagina ordinarily are extensions from infections of the external organs, but there are four types of vaginitis which account for a high proportion of gynecologic disease. All forms are marked by pruritus (itching), inflammation of variable degree, and often display a discharge, or leucorrhea, of some type. Gonorrheal vulvovaginitis is seen almost always in children, trichomonal vaginitis results from a protozoan infection, and monilial vaginitis is a fungus infection similar to thrush. Senile vaginitis occurs in elderly women following chronic irritation of the thinned, atrophic mucosa *See* CANDIDIASIS; GONORRHEA; TRICHOMONIASIS.

The most important primary lesion of the vagina is carcinoma but its incidence is low. However, extension of the much more common cervical carcinoma to the vagina occurs in advanced cases. In both types, the prognosis is dependent largely upon time of diagnosis and extent of spread. *See* ONCOLOGY.

Infrequent benign tumors include papillomas, fibromas, and hemangiomas of minor significance

Trauma of the vagina is common. Forceful coitus, including rape, accidental injury as a result of falling on penetrating objects, and the injuries resulting from childbirth may produce severe lesions and complications such as infection or scarring. Foreign materials introduced either purposely or accidentally are frequently encountered and are often bizarre in nature Certain chemicals used for douches or suppositories may cause irritation, sensitivity, or even ulcerations and chemical burns.

Vaginismus is primarily a functional disorder marked by spasmodic, painful and involuntary contraction of the lower vaginal muscles Although usually psychic in origin, it may result from remedial physical causes Intercourse is painful or impossible and often intense sensations are elicited by the lightest touch Vaginismus is not uncommonly associated with frigidity and in both cases psychotherapy may be helpful *See* PSYCHOTHERAPY [F C ST.]

Valence

A term commonly used by chemists to characterize the combining power of an element for other elements, as measured by the [illegible] f bonds to

other atoms which one atom of the *given* element forms upon chemical combination. The term also has come to signify the theory of all the physical and chemical properties of molecules that specially depend on molecular electronic structure

Thus, in water, H_2O or

$$\begin{array}{l} \mathrm{H} \\ | \\ \mathrm{O—H} \end{array}$$

the valence of each hydrogen atom is 1, the valence of oxygen 2 In methane, CH_4 or

$$\begin{array}{ccc} & \mathrm{H} & \\ & | & \\ \mathrm{H}— & \mathrm{C} & —\mathrm{H} \\ & | & \\ & \mathrm{H} & \end{array}$$

the valence of hydrogen again is 1, of carbon 4. In NaCl and CCl_4 the valence of chlorine is 1, and in CH_2 the valence of carbon is 2

Much more is known about a water molecule than that it contains two hydrogen atoms and one oxygen atom. Each OH distance is 0.957×10^{-8} cm and the HOH bond angle is 104°27′. The oxygen and hydrogen ends of the molecule are negatively and positively charged, giving it a dipole moment 1.84×10^{-18} electrostatic units (esu) The molecule absorbs infrared light strongly but is transparent to visible light Scientists are striving for an understanding of these properties and many more in terms of the fundamental theory of valence

Here valency is a synonym for valence

Combining power of an element. By the 1920s the most important facts about atoms had been established experimentally A neutral atom of atomic number Z comprises a massive nucleus of charge $+Ze$ and Z very light electrons each of charge $-e$, where $e = 4.80 \times 10^{-10}$ esu, most of the space within the atom is empty Atomic nuclei are immutable through ordinary chemical changes, when one molecule of H_2 combines with one molecule of Cl_2 to give two molecules of HCl the four nuclei (two hydrogen nuclei, or protons, of charge $+1e$ and two chlorine nuclei of charge $+17e$) are unchanged It is redistribution of electrons between atoms which constitutes chemical combination This is what valences of atoms control and this is what a theory of valence must explain

Atomic structure To understand molecule formation, then one first must understand the electronic structure of atoms According to Neils Bohr electrons in an atom move in orbits much like the orbits of planets about a sun, held to the nucleus by electrical *attractions* for it, prevented from falling into it by *centrifugal forces* A special quantum effect is operative at the atomic level, however, which possesses no analogy in the motions of planets, not all *orbits* are possible for an electron, but only those for which the *angular momentum* of the electron as it moves about the nucleus is an integral multiple of $h/2\pi$, where $h = 6.63 \times 10^{-27}$ erg sec is Planck's constant, and for which the energy is similarly quantized Furthermore, not more than two electrons can move in one orbit at once. *See* ATOMIC STRUCTURE AND SPECTRA; QUANTUM CHEMISTRY.

When the consequences of these ideas are worked out, there actually emerges the periodic classification of the elements. To cover just part of the periodic table, occupation of orbits by electrons in the lighter atoms are shown in the table where the

Electron configurations of some atoms

Atom		Orbit					
		K shell	*L* shell		*M* shell		
	Z	1s	2s	2p	3s	3p	3d
H	1	1	0	0	0	0	0
He	2	2	0	0	0	0	0
Li	3	2	1	0	0	0	0
Be	4	2	2	0	0	0	0
B	5	2	2	1	0	0	0
C	6	2	2	2	0	0	0
N	7	2	2	3	0	0	0
O	8	2	2	4	0	0	0
F	9	2	2	5	0	0	0
Ne	10	2	2	6	0	0	0
Na	11	2	2	6	1	0	0
Mg	12	2	2	6	2	0	0
Al	13	2	2	6	2	1	0

symbol $2p$ stands for three distinct orbits of the same energy and shape but differently oriented in space The lowest energy orbit is 1s, forming the K shell Next in energy are 2s and 2p, making up the L shell The 3s state is still higher, in the M shell The chemically inert gases helium, He, and neon, Ne, are characterized by closed shells of 2, and $2 + 8 = 10$ electrons respectively The next inert gas is argon, Ar, with a closed shell of $2 + 8 + 8 = 18$ electrons, followed by krypton, Kr, with $2 + 8 + 18 + 8 = 36$ electrons, and the others. *See* PERIODIC TABLE

Rule of eight. Most of the simple facts of valence (though certainly not all) follow from the postulate that atoms combine in such a way as to seek closed shell or inert gas structures (rule of eight) by the transfer of electrons between them or the sharing of a pair of electrons between them Following G. N Lewis, many molecular structures may be obtained by inspection using these rules Letting a dot represent an electron,

$$\begin{array}{c} \mathrm{H} \\ \ddot{} \\ :\ddot{\mathrm{O}}:\mathrm{H} \end{array} \qquad \begin{array}{c} \mathrm{H} \\ \ddot{} \\ \mathrm{H}:\ddot{\mathrm{C}}:\mathrm{H} \\ \mathrm{H} \end{array} \qquad [\mathrm{Na}]^{+} \qquad \left[:\ddot{\underset{..}{\mathrm{F}}}:\right]^{-}$$

In these electron-dot symbols, the electrons in the K shell are not included for atoms after He, nor are the electrons in the K and L shells for atoms following Ne

Hydrogen has a valence of 1, because one more electron will give a hydrogen atom an inert gas structure Carbon can form four bonds because four more electrons give it the neon electronic structure

Bond types The bond between two atoms is covalent if one electron in the *bonding* electron

pair comes from each atom, as in H:H or the CH bonds in CH_4. It is coordinate covalent if both electrons come from one atom, as the boron-nitrogen bond in the compound

$$\begin{array}{c} \;\;\ddot{:F}:H \\ :\ddot{F}:\ddot{B}:\ddot{N}:H \\ \;\;\ddot{:F}:H \end{array}$$

If there is complete transfer of electrons from one atom to another the bond is electrovalent or ionic, as in sodium fluoride, NaF Bonds intermediate in type are possible; the bond in hydrogen fluoride, HF, is between covalent and ionic. An ionic bond X^+Y^- will be more stable the less the ionization potential of X and the greater the affinity of Y for electrons, that is, when X is a metallic element from the lower left corner of the periodic table and Y is a nonmetallic element from the upper right corner. Bond type can be inferred from both chemical and physical evidence. *See* CHEMICAL BINDING, ELECTRONEGATIVITY.

Bonds involving one or three electrons are known, but they are rare, H_2^+ and HeH are examples Multiple bonds between atoms are common and important; examples are the carbon-carbon bond and the carbon-oxygen bonds in ethylene and carbon dioxide

$$H_2C::CH_2 \quad \text{or} \quad H_2C{=}CH_2$$

and

$$O::C::O \quad \text{or} \quad O{=}C{=}O$$

For discussion of a bond of special importance in biology, *see* HYDROGEN BOND.

Valence electrons are the electrons of an atom that can participate in chemical binding, for example, for H and He the 1*s* electrons, for Li through Ne the 2*s* and 2*p* electrons, and for Na the 3*s* electron

Oxidation-reduction. As generally used and here defined, the word valence is ambiguous. Before a value can be assigned to the valence of an atom in a molecule, the electronic structure of the molecule must be exactly known, and this structure must be describable simply in terms of simple bonds. In practice neither of these conditions is ever precisely fulfilled A term not so ambiguous is oxidation number or valence number. Oxidation numbers are useful for the balancing of oxidation-reduction equations, but they are not related simply to ordinary valences Thus the valence of carbon in CH_4, $CHCl_3$, and CCl_4 is 4, oxidation numbers of carbon in these three substances are -4, $+2$, and $+4$. *See* OXIDATION-REDUCTION

Quantum theory of valence. The above theory of valence is inadequate in at least three ways. First, it fails to account for many experimental facts, such as why the six C—C bonds in the molecule benzene, C_6H_6, are physically and chemically equivalent, what the electronic structures of the boron hydrides are, why the H—H bond is much stronger than the C—C bond, why CO_2 is a linear molecule but H_2O nonlinear, and what principles govern the rates of chemical combination Secondly, the explanations that are offered are not physically satisfying. The stability conferred upon a molecule by the sharing of a pair of electrons by two atoms is established, but what is the real origin of this stability? And thirdly, the theory is not comprehensive or quantitative enough to allow correlation and prediction of the many different properties of molecules. Dozens of properties of molecules can be measured, many to high accuracy. A theory should ultimately, and quantitatively, account for all of these. *See* MOLECULAR STRUCTURE AND SPECTRA.

The quantum theory of valence does not have these faults It is based on the new precise laws of physics for the atomic domain which were formulated in the 1920s by E. Schrödinger and others, the discipline called quantum mechanics. The quantum ideas of M. Planck and N. Bohr require modification to take care of experimental observations that electrons and other particles at times act like waves Like waves, they interfere when they are on top of one another in a manner that can be precisely calculated. According to nineteenth-century physics, an electron moving about a proton would collapse onto it In the Bohr theory this collapse is prevented by a special quantum hypothesis; in the new mechanics it is prevented by elementary energy considerations. It would be favored by the attractive potential energy of the particle pair, but it turns out to be catastrophic for their kinetic energy Instead of collapse a compromise is reached; the electron, or wave, is smudged out over a region about the nucleus which defines the atomic size *See* QUANTUM MECHANICS; UNCERTAINTY PRINCIPLE

The pattern of the periodic table comes out as before. The orbits of Bohr are replaced by new entities, orbitals, which represent not the paths of the electrons but the amplitudes of the electron waves at different points in space. Furthermore, electrons are treated as if they were spinning, but only in two possible ways The rule that generates the periodic table then is that in an atom no two electrons can occupy the same atomic orbital with the same spin *See* EXCLUSION PRINCIPLE.

In a chemical bond again there is interplay of kinetic and potential energies An electron pair will tend to be shared by two atoms instead of being located on one of them if that situation is energetically favorable. The region between nuclei is more favorable for the potential energy of electron-nuclear attraction than other regions the same distance from just one nucleus Moving in this restricted region is not as favorable for the kinetic energy as moving on individual atoms, but the potential energy predominates when a bond is formed The normal covalent bond may be described as two electrons occupying one molecular orbital, rather than two distinct atomic orbitals, with opposite spins because the exclusion principle is still operative.

When a detailed examination is made of these effects with the new theory, the stabilities of actual molecules and other of their properties can be quantitatively accounted for. In particular, if two atoms approach which have low-energy atomic orbitals which overlap each other in space, and if two electrons are available, the conditions are favorable for forming, with evolution of heat, a chemical bond. It follows that the valence of an atom is given by the number of unpaired electrons it possesses, an old basic rule of valence.

The greater the overlap between two atomic orbitals, the stronger the bond that can be formed with them (criterion of maximum overlapping). This condition may be regarded as determining the shapes of molecules. Two or more orbitals of comparable energy, as 2s and 2p orbitals, can be combined (hybridized) to give orbitals concentrated along certain directions in space, and these are the orbitals that participate in directed bond formation. In the carbon atom, for instance, the four electrons in the 2s and 2p subshells are potential valence electrons. The two 2s electrons are paired, however, so that to make four bonds possible one of these must be promoted to a vacant 2p orbital. Four bonds then are possible, in various directions. Four equivalent bonds can be formed, tetrahedrally directed, as in CH_4. Three bonds in a plane and one other less strong one can be formed, as in $CH_2{=}CH_2$. In this manner Linus Pauling and others have accounted for a multitude of phenomena in stereochemistry.

The peculiar binding in benzene and other aromatic molecules has been explained, together with its consequences for chemical reactivity. The principles governing reaction rates have been formulated and applied.

Research in valence theory through the 1930s and 1940s has led to understanding of a great deal of chemistry, and it has contributed toward acceptance of the language of modern physics as a proper language for chemistry. However, considerable research in this field continues. New substances with new types of bonds are being synthesized constantly (for example ferrocene, the first "molecular sandwich," in 1951). New physical methods for studying molecules are constantly revealing more intimate details of molecular structure which demand explanation (for example, the new techniques of magnetic resonance). Also, intensive work continues with applications of large digital electronic computing machines to problems of valence. It is likely that accurate determination of many properties of molecules containing only light atoms will be achieved by such computational methods in the 1960s. See CHELATION; CHEMICAL STRUCTURES, COORDINATION CHEMISTRY; KINETICS (CHEMICAL); MAGNETIC RESONANCE; ORGANIC CHEMISTRY; ORGANOMETALLIC COMPOUND, RESONANCE (MOLECULAR STRUCTURE), STEREOCHEMISTRY [R G P]

Bibliography: C. A. Coulson, *Valence*, 1952; L. Pauling, *General Chemistry*, 2d ed., 1956; L. Pauling, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals*, 3d ed., 1959; F. O. Rice and E. Teller, *The Structure of Matter*, 1949.

Valence band

The highest electronic energy band in a semiconductor or insulator which can be filled with electrons. The electrons in the valence band correspond to the valence electrons of the constituent atoms. In a semiconductor or insulator, at sufficiently low temperatures, the valence band is completely filled and the conduction band is empty of electrons (see CONDUCTION BAND). Some of the high energy levels in the valence band may become vacant as a result of thermal excitation of electrons to higher energy bands or as a result of the presence of impurities. When some electrons are missing, the remaining ones may be redistributed among the energy levels within the valence band under an applied electric field, giving rise to an electric current. The net effect of the valence band is then equivalent to that of a few particles which are equal in number and similar in motion to the missing electrons but each of which carries a positive electronic charge. These "particles" are referred to as holes. See HOLES IN SOLIDS, see also BAND THEORY OF SOLIDS; INSULATOR ELECTRIC, SEMICONDUCTOR [H Y F]

Valine


Valine

Physical constants of the L isomer at 25 C
pK_1 (COOH) 2.32, pK_2 (NH_3^+) 9.62
Isoelectric point 5.96
Optical rotation $[\alpha]_D(H_2O)$ +5.6
$[\alpha]_D$(5 N HCl) +28.3
Solubility (g/100 ml H_2O) 8.85

An amino acid considered essential for normal growth of animals. The amino acids are characterized physically by the following: (1) the pK_1, or the dissociation constant of the various titratable groups; (2) the isoelectric point, or pH at which a dipolar ion does not migrate in an electric field; (3) the optical rotation or the rotation imparted to a beam of plane-polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml, (4) solubility. See EQUILIBRIUM, IONIC; ISOELECTRIC POINT, OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS

The biosynthetic precursor as well as the deamination product of valine, is α ketoisovaleric acid. It is also a precursor of leucine and of the pantoic acid moiety of pantothenic acid.

Valine is biosynthesized from pyruvic acid (see AMINO ACIDS). Most or all of the enzymes concerned also catalyze the analogous reac-

pair comes from each atom, as in H:H or the CH bonds in CH_4. It is coordinate covalent if both electrons come from one atom, as the boron-nitrogen bond in the compound

$$\begin{array}{ccccc} & & \cdot\cdot & & \\ & & :F: & H & \\ \cdot\cdot & & \cdot\cdot & \cdot\cdot & \\ :F: & B & : & N: & H \\ \cdot\cdot & & \cdot\cdot & \cdot\cdot & \\ & & :F: & H & \\ & & \cdot\cdot & & \end{array}$$

If there is complete transfer of electrons from one atom to another the bond is electrovalent or ionic, as in sodium fluoride, NaF. Bonds intermediate in type are possible; the bond in hydrogen fluoride, HF, is between covalent and ionic. An ionic bond X^+Y^- will be more stable the less the ionization potential of X and the greater the affinity of Y for electrons, that is, when X is a metallic element from the lower left corner of the periodic table and Y is a nonmetallic element from the upper right corner. Bond type can be inferred from both chemical and physical evidence. *See* CHEMICAL BINDING, ELECTRONEGATIVITY.

Bonds involving one or three electrons are known, but they are rare; H_2^+ and HeH are examples. Multiple bonds between atoms are common and important; examples are the carbon-carbon bond and the carbon-oxygen bonds in ethylene and carbon dioxide

$$H_2C::CH_2 \quad \text{or} \quad H_2C{=}CH_2$$

and $$O::C::O \quad \text{or} \quad O{=}C{=}O$$

For discussion of a bond of special importance in biology, *see* HYDROGEN BOND.

Valence electrons are the electrons of an atom that can participate in chemical binding, for example, for H and He the 1*s* electrons, for Li through Ne the 2*s* and 2*p* electrons, and for Na the 3*s* electron.

Oxidation-reduction. As generally used and here defined, the word valence is ambiguous. Before a value can be assigned to the valence of an atom in a molecule, the electronic structure of the molecule must be exactly known, and this structure must be describable simply in terms of simple bonds. In practice neither of these conditions is ever precisely fulfilled. A term not so ambiguous is oxidation number or valence number. Oxidation numbers are useful for the balancing of oxidation-reduction equations, but they are not related simply to ordinary valences. Thus the valence of carbon in CH_4, $CHCl_3$, and CCl_4 is 4; oxidation numbers of carbon in these three substances are −4, +2, and +4. *See* OXIDATION-REDUCTION.

Quantum theory of valence. The above theory of valence is inadequate in at least three ways. First, it fails to account for many experimental facts, such as why the six C—C bonds in the molecule benzene, C_6H_6, are physically and chemically equivalent, what the electronic structures of the boron hydrides are, why the H—H bond is much stronger than the C—C bond, why CO_2 is a linear molecule but H_2O nonlinear, and what principles govern the rates of chemical combination. Secondly, the explanations that are offered are not physically satisfying. The stability conferred upon a molecule by the sharing of a pair of electrons by two atoms is established, but what is the real origin of this stability? And thirdly, the theory is not comprehensive or quantitative enough to allow correlation and prediction of the many different properties of molecules. Dozens of properties of molecules can be measured, many to high accuracy. A theory should ultimately, and quantitatively, account for all of these. *See* MOLECULAR STRUCTURE AND SPECTRA.

The quantum theory of valence does not have these faults. It is based on the new precise laws of physics for the atomic domain which were formulated in the 1920s by E. Schrödinger and others, the discipline called quantum mechanics. The quantum ideas of M. Planck and N. Bohr require modification to take care of experimental observations that electrons and other particles at times act like waves. Like waves, they interfere when they are on top of one another in a manner that can be precisely calculated. According to nineteenth-century physics, an electron moving about a proton would collapse onto it. In the Bohr theory this collapse is prevented by a special quantum hypothesis; in the new mechanics it is prevented by elementary energy considerations. It would be favored by the attractive potential energy of the particle pair, but it turns out to be catastrophic for their kinetic energy. Instead of collapse a compromise is reached: the electron, or wave, is smudged out over a region about the nucleus which defines the atomic size. *See* QUANTUM MECHANICS, UNCERTAINTY PRINCIPLE.

The pattern of the periodic table comes out as before. The orbits of Bohr are replaced by new entities, orbitals, which represent not the paths of the electrons but the amplitudes of the electron waves at different points in space. Furthermore, electrons are treated as if they were spinning, but only in two possible ways. The rule that generates the periodic table then is that in an atom no two electrons can occupy the same atomic orbital with the same spin. *See* EXCLUSION PRINCIPLE.

In a chemical bond again there is interplay of kinetic and potential energies. An electron pair will tend to be shared by two atoms instead of being located on one of them if that situation is energetically favorable. The region between nuclei is more favorable for the potential energy of electron-nuclear attraction than other regions the same distance from just one nucleus. Moving in this restricted region is not as favorable for the kinetic energy as moving on individual atoms, but the potential energy predominates when a bond is formed. The normal covalent bond may be described as two electrons occupying one molecular orbital, rather than two distinct atomic orbitals with opposite spins because the exclusion principle is still operative.

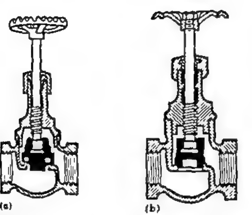


Fig 2 Globe valves (a) With gasket in disk. (b) With ground metal-faced disk.

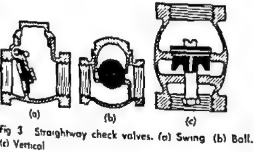


Fig 3 Straightway check valves. (a) Swing (b) Ball. (c) Vertical

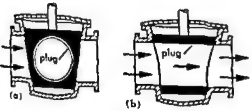


Fig 4 Plug valve (a) Closed (b) Open

tions of public authorities and insurance underwriters. They must open automatically when the pressure exceeds a predetermined value; they must allow the pressure to drop a predetermined amount before closing to avoid chattering, instability, and damage to the valve and the valve seat; they must have adjustment features for both the relieving and blowdown pressures; and they must be tamperproof after setting by responsible licensed operators. See SAFETY VALVE.

Hydraulic turbine valves. For hydraulic turbines and hydroelectric systems, valves and gates control water flow for (1) regulation of power output at sustained efficiency and with minimum wastage of water and (2) safety under the inertial flow conditions of large masses of water. Valve sizes are usually large so that power operation is necessary. Carefully streamlined construction to minimize fluid dynamic losses must be accompanied by ample provision to withstand shock and damaging effects of hydraulic inertia (see WATER HAMMER). Gate, butterfly (Fig 5), telescoping, and needle constructions (Fig 6) are variously employed. Wicket or cylinder gates regulate the flow of water to a reaction turbine at the speed ring while a governor-operated needle valve regulates flow to a Pelton impulse unit.

Steam-engine valves. To control the kinematics of the cycle, steam-engine valves range from simple D-slide and piston valves to multiported types. Slide valves control admission and release of steam to and from a double-acting cylinder by a single moving valve mechanism giving the necessary lap, lead, and angle of advance to accomplish the predetermined values of cutoff and compression. Multiported valves such as plug, Corliss, or poppet valves provide four valves for a double-acting cylinder. Each valve serves a single purpose of admission or exhaust for the head or crank end (see STEAM ENGINE). The uniflow construction uses a poppet valve for admission with a row of exhaust ports alternately covered and uncovered by the engine piston. Plug and slide valves are limited to low pressures and temperatures (200 psi and 100°F of superheat); poppet valves will operate on the maximum pressures and temperatures of steam engine practice without warping or leaking. Many types of reversing gear have been perfected which use the same slide valve or piston valve for both forward and backward rotation of an engine, as in railroad and marine service.

Internal combustion engine valves. Poppet valves are used almost exclusively in internal-combustion reciprocating engines because of the demands for tightness with high operating pressures and temperatures. The valves are generally 2 in. in diameter or smaller on high speed automotive-type engines, are cam-operated, spring loaded, with their lift a small fraction of an inch, and gas veloc-

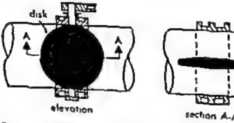


Fig 5 Butterfly-type valve for penstock is typically 25 ft in diameter

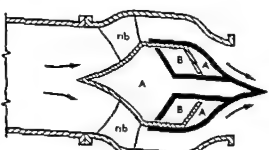
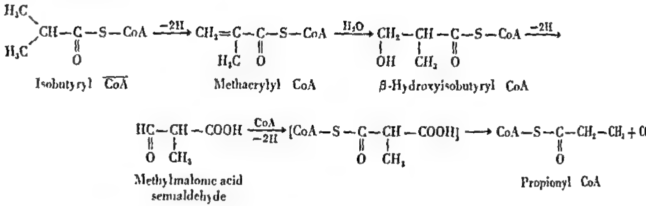


Fig 6 Large needle valve is actuated hydraulically by pressure in chambers A to close and in annular chamber B to open.

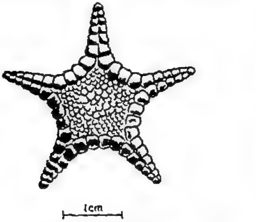


Final steps in metabolic degradation.

form isobutyrylcoenzyme A (isobutyryl-CoA) The final steps in metabolic degradation (see equation) are the conversion of isobutyrvl-CoA to CO_2 and propionyl-CoA by the formation of the following compounds: methacrylyl-CoA, β-hydroxyisobutyryl-CoA, and methylmalonic acid semialdehyde [E.A.AD.]

Valvata

A suborder of Phanerozonia in which the upper marginals lie directly over, and not alternate with, the corresponding lower marginals. The tube-feet have terminal sucking disks (see illustration).



A representative valvate starfish, Iconaster perierctus (After Fisher, 1919)

Paxillae may be present or lacking The group is well represented in all existing seas from low-tide to abyssal depths. *See* PHANEROZONIA [H.B.F.]

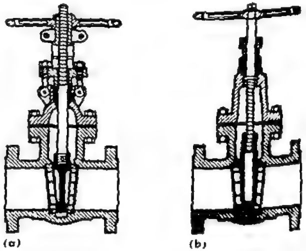
Valve

A flow-control device. This article deals with valves for fluids, liquids, and gases (for electrical valves, see ELECTRON TUBE). Valves are used to regulate the flow of fluids in piping systems and machinery. In machinery the flow phenomenon is frequently of a pulsating or intermittent character and the valve, with its associated gear, contributes a timing feature.

Pipe valves. The valves commonly used in piping systems are gate valves (Fig. 1), usually operated closed or wide open and seldom used for throttling; globe valves (Fig. 2), frequently fitted with a renewable disk and adaptable to throttling operations; check valves (Fig. 3), for automatically limiting flow in a piping system to a single direction; and plug cocks (Fig 4), for operation in the open or closed position by turning the plug through 90° and with a shearing action to clear foreign matter from the seat.

Valves may have various structural features such as outside stem and yoke; packless construction; angle, as opposed to straightway flow; power instead of manual operation; and combined nonreturn and stop-valve arrangements. Valves are made in a wide assortment of materials, and a wide variety of trim, with brass or bronze for general service, cast iron for low steam pressures and temperatures (less than 250 psi) and for hydraulic pressures below 800 psi; steel and alloy steels for the highest operating pressures and temperatures (such as 5000 psi, 1200°F steam); and selected metals for chemical and process applications Most valves are manufactured and available as hardware and comply with the requirements of the ASTM, ASA, and ASME as to material and dimensional standards. They are variously offered as flanged, screwed, welded, sweated, or compression fitted for connection to pipe, machinery, and fittings.

Safety and relief valves are automatic protective devices for the relief of excess pressure They are usually rigorously specified under the legal regula



(a) (b)

Fig 1 Gate valves with disk gates shown solid. (a) Rising threaded stem shows when ... is open. (b) Nonrising stem valve requ...

held closed by a compressed helical spring. The spring must be strong enough to overcome the inertia of the valve mechanism and to hold the valve follower on the cam at all times. The valve is opened wide by lifting it from its seat a distance approximately equal to 25% of the valve diameter At this lift the cylindrical opening between valve and seat will be about as large as the cross section of the flow passage to the valve seat.

Valve action requires that the valve be streamlined and as large as possible to give maximum flow, yet be of low inertia, so that it follows the prescribed motion at high engine speed. Valves are usually made of stainless, nonscaling alloy which will keep its strength and shape at high temperature The exhaust valve is sometimes made hollow and partially filled with metallic sodium. In operation the sodium melts and shakes back and forth, transferring heat from the valve head to the stem and valve guide Valve seat inserts of tough heat-resistant alloys are used in cylinders of high performance engines or when the cylinder material alone is not strong enough for this purpose.

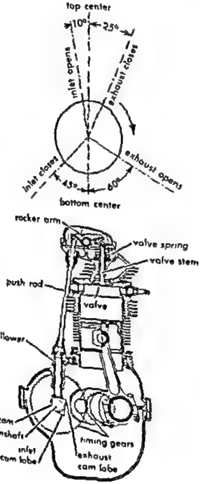


Fig 1 Cross section of I-head cylinder and valve timing diagram of a four-cycle engine (From C H Chatfield, C F. Taylor, and S Ober, *The Airplane and Its Engine*, 5th ed., McGraw-Hill, 1949)

Cam action. Engine valves are usually opened by means of cams Riding on each cam is a follower, or valve lifter, which may be a flat or slightly convex surface, or a roller (Fig. 2). The valve is opened by forces applied to the end of the valve stem through a mechanical linkage actuated by the cam follower. In overhead valve engines, the cam shaft may be mounted on the cylinder head near the valves. In less expensive construction, the cam shaft

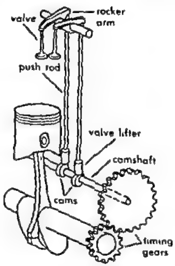


Fig 2 Isometric of moving parts of valve mechanism (*Buick Motor Division of General Motors Corp*)

is placed in the crank case. The operating linkage then consists of cam follower, push rod, and rocker arm. The push rod is a light rod or tube with ball ends, which carries the motion of the cam follower to the rocker arm. The rocker arm is a lever pivoted near its center so that, as the push rod raises one end, the other end depresses the valve stem, opening the valve.

To insure tight closing of the valve even when the valve stem lengthens from thermal expansion, the valve train is adjusted to provide some clearance when the follower is on the low part of the cam (see HYDRAULIC VALVE LIFTER). The cam shape includes a ramp which reduces shock by starting the lift at about two feet per second even though the clearance varies from time to time. To open the valve quickly, an acceleration of about 400 times gravity (for automotive-size engines) is used Excessive acceleration deflects the valve linkage, giving false motion to the valve, causing it to close at high velocity before the closing ramp is reached The cam surface between the opening and closing acceleration sections includes the point of maximum lift and maximum deceleration If this part of the cam is sinusoidal and the valve spring is properly designed the spring forces remain proportional to the forces required by the cam for the deceleration of the valve The highest operating speed for a given lift can thus be obtained without the follower leaving the cam Sometimes the valve is held at maximum lift for a time This part of the cam is called the dwell

ity of 200–300 ft/sec. They are cooled by transfer of heat to the engine jacket mostly through the valve stem. In heavy-duty units, the stem may have

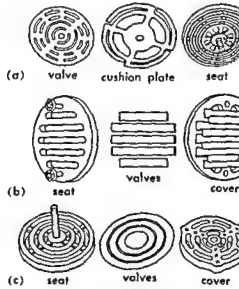


Fig 7 Air and gas compressor valves *(a)* Ingersoll-Rand ring plate. *(b)* Worthington feather valve. *(c)* Chicago pneumatic simplate valve *(From T Baumeister, Marks' Mechanical Engineers' Handbook, McGraw-Hill, 6th ed , 1958)*

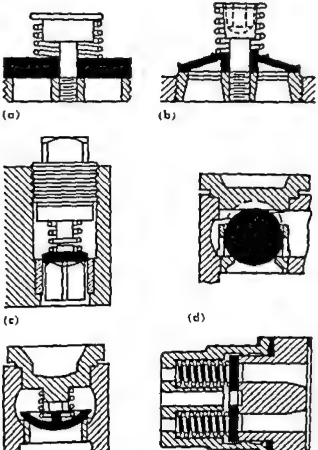


Fig. 8 Pump valves *(a, b)* Disk valves are for moderate pressure. *(c)* Conical-faced wing valves are for high pressure (1000 psi ±) *(d, e)* Ball and rounded valves are for viscous liquids. *(f)* Double-ported valves are for high-speed reciprocating pump.

a partial mercury or sodium filling which aids, by inertia, in conducting heat to the valve stem and guides. Exhaust valves are subject to the extremes of temperature effects and must accordingly be most carefully designed and constructed of alloy materials. *See* CAM MECHANISM; VALVE TRAIN

Compressor valves. In compressors, valves are usually automatic, operating by pressure difference (<5 psi) on the two sides of a movable, spring loaded member and without any mechanical linkage to the moving parts of the compressor mechanism (Fig. 7). The practical objectives are timing, tightness, and life. Pressure drop should be a minimum. Moving parts must be light but spring loadings should be heavy enough to avoid chatter and to give certainty to the valve action. Air speeds are in the order of 5000 ft/min through the valves. Valves are frequently made interchangeable for inlet and exhaust service in the interest of reduced investment. On high-pressure (>1000 psi) service, automatic poppet valves are usually substituted for the more common plate valves to give greater strength and tightness under loading *See* COMPRESSOR.

Pump valves. Like those for compressors, pump valves are usually of the automatic type operating by pressure difference The service conditions, however, are very dissimilar because of the noncompressibility of liquids; the presence of entrained solids like grit, fibers, and sludge, the corrosive potential of chemicals like acids and alkalies; the high viscosity of many liquids; the vapor pressure at the pump-operating temperature; and inertia effects accompanying discontinuity of the liquid column. Fluid speeds are low (200–300 ft/min with cold water and 100 ft/min with viscous liquids). A multiplicity of small valves (4 in. ± in diameter) arranged in decks with a positive suction is a common construction. Various constructions are used (Fig 8) [T.B.]

Bibliography. T Baumeister, *Marks' Mechanical Engineers' Handbook*, 6th ed , 1958.

Valve train

The valves and valve-operating mechanism by which an internal-combustion engine takes air or fuel-air mixture into the cylinders and discharges combustion products to the exhaust

Mechanically an internal-combustion engine is a reciprocating pump, able to draw in a certain amount of air per minute Since the fuel takes up little space but needs air with which to combine, the power output of an engine is limited by its air-pumping capacity. It is essential that the flow through the engine be restricted as little as possible This is the first requirement for valves. The second is that they close off the cylinder firmly during the fuel combustion and power strokes.

Valve action. In most four-stroke engines the valves are of the inward opening poppet type, with valve head ground to fit a conical seat in the cylinder block or cylinder head (Fig. 1). The valve head is held concentric with its seat by a cylindrical stem running in a valv[illegible] he valve is

fourth segment of the body. Adult females often have the young attached to the antennae, which thus acquire a velvety appearance. Some Arcturidae are found inshore on algae, on hydroids, or even amongst the spines of sea urchins, but more species come from deeper water, which no doubt accounts for our incomplete knowledge of the group [E N.]

Bibliography· J. Kjennerud, Ecological observations on *Idotea neglecta* G. O. Sars, *Univ. Bergen Årbok Naturvitenskap Rekke*, 7:1–47, 1950; A P. M Lockwood and P. C. Croghan, The chloride regulation of the brackish and freshwater races of *Mesidotea entomon* (L.), *J Exptl Biol*, 34(2): 253–258, 1957; E. Naylor, The ecological distribution of British species of *Idotea* (Isopoda), *J. Animal Ecol.*, 24(2):255–269, 1955; E Naylor, The occurrence of *Idotea metallica* Bosc. in British waters, *J. Marine Biol. Assoc. United Kingdom*, 36(3):599–602, 1957, H Richardson *A Monograph on the Isopods of North America*, U.S. Natl Museum Bull 54, 1905, G O. Sars, *An Account of the Crustacea of Norway*, vol. 2, 1899

Vampyromorpha

An order of dibranchiate cephalopod mollusks *Vampyroteuthis infernalis*, an inhabitant of the deeper waters of tropical and temperate seas, is the only representative of this order It is considered to be a link form between the decapod and octopod mollusks *See* CEPHALOPODA. [C.B C]

Van Allen radiation

A high intensity of charged particles in immense belts in the space surrounding the Earth These belts were discovered with early *satellites* of the United States, *Explorer* I (*Satellite 1958α*) and *Explorer* III (Satellite 1958γ), and *reported* by a scientific team at the State University of Iowa in May, 1958 Detailed studies have continued by this group, by IGY workers in the Soviet Union, and by others. *Argus* experiments of August–September, 1958 furthered knowledge *of the belts* by artificial injection of β-decay *electrons* from the fission fragments of high altitude detonations of small-yield atomic *devices* and their subsequent observation by *Explorer* IV (Satellite 1958ε), by sounding rockets, and by *observation* of other geophysical effects. The Argus experiments were proposed by N. C. Christofilos

Intensity of radiation. The charged particles are *temporarily trapped* in the geomagnetic field, thereby *producing* an intensity distribution as illustrated The diagram *is* a geomagnetic meridian section of a three-dimensional *figure* of revolution around the geomagnetic axis. Locations of the two principal *zones or belts* are derived from Geiger-tube observations Plotted intensities are in true counting rates *of a counter* of specified characteristics. Throughout the hatched regions the rate exceeds 10,000 counts/sec, the *maximum* being about 25,000 The structure of the outer belt fluctuates widely with time.

It is probable that the outer radiation zone, in particular, plays a central role *in many* geophysical phenomena: auroras, geomagnetic storms, airglow, and atmospheric heating Study of detailed mechanisms is being pursued by many investigators. The discovery, by S. Dolginov and N. Pushkov, of the terrestrial ring current in the inner portion of the outer radiation zone is of special note.

Data from *Pioneer* IV of March 3–6, 1959, confirmed the general structure of the region of trapped radiation but showed a great *enhancement* of intensity and a considerable extension of the radial limits of the outer zone, following the great *M*-region event on the Sun on February 25, 1959. In addition, these data provide information on the absorptivity of the radiation in both inner and outer zones and give a new determination of the cosmic-ray intensity in interplanetary space.

Nature of radiation. The trapped *radiation* consists of protons and electrons constrained to bound orbits by the geomagnetic field in the manner visualized in classical theoretical studies by H Poincaré, C Stoermer, and H. Alfvén.

Radiation in the inner zone is quite different in composition from that in the outer zone. In the heart of the inner zone at an altitude of about 3600 km on the geomagnetic equator, the approximate *composition includes* electrons of energy greater than 40 kev having a maximum of unidirectional intensity of $\sim 2 \times 10^9/(\mathrm{cm}^2)(\mathrm{sec})$(steradian); electrons of energy greater than 600 kev which have a maximum of unidirectional intensity of $\sim 1 \times 10^7/(\mathrm{cm}^2)(\mathrm{sec})$(steradian); and protons of energy greater than 40 Mev which have an *omnidirectional* intensity of $\sim 2 \times 10^4/(\mathrm{cm}^2)(\mathrm{sec})$. The inner zone is observed to be relatively free of temporal fluctuations.

The outer zone, as mentioned earlier, undergoes marked temporal fluctuations. It appears that *Pioneer IV passed* through the outer zone during a *period of exceptional* enhancement (March 3, 1959). The *composition* of the radiation in the heart of the outer zone (altitude about 20,000 km on the geomagnetic *equator*) *on* that date was approximately as follows: (1) electrons of energy *greater than* 40 kev having an omnidirectional intensity of $\sim 1 \times 10^{11}/(\mathrm{cm}^2)(\mathrm{sec})$, (2) electrons of energy greater than 200 kev having an omnidirectional intensity less than $1 \times 10^9/(\mathrm{cm}^2)(\mathrm{sec})$, and (3) *protons* of energy greater than 60 Mev having an omnidirectional intensity less than $1 \times 10^2/(\mathrm{cm}^2)(\mathrm{sec})$. There *was no* significant information for protons of en-

[illegible]

w appears

the *inner*

zone are the products of the radioactive decay of neutrons moving outward from the top *of the atmosphere*. These neutrons are released from cosmic ray-induced *nuclear* disintegrations in the atmosphere.

The outer zone almost certainly owes its existence *to ionized* solar gas *ejected* from the Sun. This gas may travel to the Earth and be injected into the geomagnetic *field* Presumably, the outer zone contains large *densities* of protons and elec-

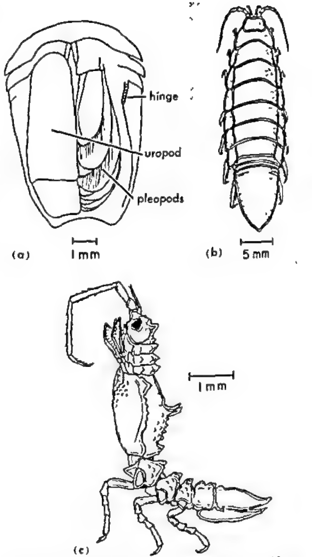
Timing. At low piston speeds gas friction and inertia are small and the valves should be opened or closed as the piston reaches the end of the appropriate stroke. At high piston speeds the inertia of the charge in the inlet passages causes the gas to continue to flow into the cylinder long after the piston has started up on the compression stroke. To trap the maximum amount of fresh charge in the cylinder it is therefore necessary to delay the closure of the inlet valve up to about 60° after bottom center, depending upon the geometry of the inlet passages and the piston speed. Late inlet closing increases the torque or pulling power of the engine at high speeds, but reduces torque at low speeds, because of the loss of fresh charge back into the inlet system during compression. To reduce the work required to expel the products of combustion during the exhaust stroke, it is necessary, at high speeds, to open the exhaust valve considerably before bottom center so that most of the gas is blown out and the cylinder pressure has fallen nearly to atmospheric before the exhaust stroke begins In supercharged engines the inlet valve often opens near top center before the exhaust valve closes The time during which both valves are open is called the overlap period Overlap permits scavenging of the combustion space with fresh charge, for valve cooling and increased power, but results in poor idling characteristics because combustion products are drawn back through the engine by the low pressure of the inlet system when the throttle is nearly closed. [A.R R]

Valvifera

A suborder of isopod crustaceans resembling woodlice and pill bugs (*see* ISOPODA). They are distinguished from these and other isopods by having a pair of flat, valvelike uropods which hinge laterally and fold inward beneath the rear part of the body. The uropods open like doors to reveal five pairs of leaflike appendages, or pleopods, which are used for respiration and swimming. Members of the group are almost exclusively marine Some species are omnivorous and have been found in dense populations on accumulations of broken algae and fish waste Many thousands are sometimes taken in a

ported as a serious pest of catches of fish retained alive in keep nets in the harbor at Bergen, Norway

The suborder is divided into two main families, the Idoteidae and the Arcturidae Idoteidae have the body more or less flattened, as is typical in isopods. The seven pairs of walking legs are all similar, though those on the posterior segments are longer and more slender Species in this family are more common inshore than in great depths and can readily be found on intertidal seaweeds or recently washed up wrack. Others may also be found in surface drift weed. Specimens of the American *Idotea metallica*, which is not a British form, are regularly found off the west coasts of Britain. They are presumably carried from the of the United



(a) The under surface of the posterior end of *Idotea*, with one uropod removed at the hinge to reveal the pleopods (*Marine Biological Assoc.*). (b) *Idotea neglecta* (c) *Arcturella*, a member of the family Arcturidae (*from G O Sars, An Account of the Crustacea of Norway, vol. 2, 1897*).

States on drifting weed and debris in the North Atlantic drift Several forms in this family are able to tolerate brackish water, and one race of the species *Mesidotea entomon* is even established in fresh-water lakes in Sweden. This species is of particular interest in the study of the manner in which fresh-water races have evolved from brackish water forms Physiological and geological evidences show that the colonization of fresh water by *Mesidotea* has taken place since the last ice age, but the method by which osmoregulation occurs has not yet been determined

Arcturidae have an almost cylindrical body and possess extremely long antennae Only the last three pairs of legs compare with those of Idoteidae; the first four pairs are fragile and without claws, and they project forward beneath the head. Many species have the habit of clinging to an object with the posterior legs and moving the front part to and fro in search of food, which is seized by the antennae. The mobility of the anterior part is increased in some species by the enormous elongation of the

fourth segment of the body Adult females often have the young attached to the antennae, which thus acquire a velvety appearance Some Arcturidae are found inshore on algae, on hydroids, or even amongst the spines of sea urchins, but more species come from deeper water, which no doubt accounts for our incomplete knowledge of the group [E N.]

Bibliography J Kjennerud, Ecological observations on *Idotea neglecta* G O Sars, *Univ. Bergen Arbok, Naturvitenskap Rekke*, 7:1–47. 1950, A P.M Lockwood and P. C. Croghan, The chloride regulation of the brackish and freshwater races of *Mesidotea entomon* (L.), *J Exptl. Biol.*, 34(2): 253-258, 1957, E Naylor, The ecological distribution of British species of *Idotea* (Isopoda), *J. Animal Ecol.*, 24(2) 255–269, 1955; E. Naylor, The occurrence of *Idotea metallica* Bosc. in British waters, *J Marine Biol Assoc. United Kingdom*, 36(3) 599–602, 1957, H Richardson *A Monograph on the Isopods of North America*, U S Natl. Museum Bull 54, 1905, G O Sars, *An Account of the Crustacea of Norway*, vol 2, 1899

Vampyromorpha

An order of dibranchiate cephalopod mollusks *Vampyroteuthis infernalis*, an inhabitant of the deeper waters of tropical and temperate seas is the only representative of this order. It is considered to be a link form between the decapod and octopod mollusks *See* CEPHALOPODA [C B C]

Van Allen radiation

A high intensity of charged particles in immense belts in the space surrounding the Earth These belts were discovered with early satellites of the United States *Explorer* I (Satellite 1958α) and *Explorer* III (Satellite 1958γ) and reported by a scientific team at the State University of Iowa in May, 1958 Detailed studies have continued by this group, by IGY workers in the Soviet Union, and by others Argus experiments of August–September, 1958 furthered knowledge of the belts by artificial injection of β-decay electrons from the fission fragments of high altitude detonations of small-yield atomic devices and their subsequent observation by *Explorer* IV (Satellite 1958ϵ), by sounding rockets, and by observation of other geophysical effects. The Argus experiments were proposed by N C Christofilos.

Intensity of radiation. The charged particles are temporarily trapped in the geomagnetic field, thereby producing an intensity distribution as illustrated The diagram is a geomagnetic meridian section of a three-dimensional figure of revolution around the geomagnetic axis Locations of the two principal zones or belts are derived from Geiger-tube observations. Plotted intensities are in true counting rates of a counter of specified characteristics. Throughout the hatched regions the rate exceeds 10 000 counts/sec, the maximum being about 25 000 The structure of the outer belt fluctuates widely with time.

It is probable that the outer radiation zone, in particular, plays a central role in many geophysical phenomena: auroras, geomagnetic storms, airglow, and atmospheric heating. Study of detailed mechanisms is being pursued by many investigators The discovery, by S. Dolginov and N. Pushkov, of the terrestrial ring current in the inner portion of the outer radiation zone is of special note.

Data from *Pioneer* IV of March 3–6, 1959, confirmed the general structure of the region of trapped radiation but showed a great enhancement of intensity and a considerable extension of the radial limits of the outer zone, following the great *M*-region event on the Sun on February 25, 1959. In addition, these data provide information on the absorptivity of the radiation in both inner and outer zones and give a new determination of the cosmic-ray intensity in interplanetary space

Nature of radiation. The trapped radiation consists of protons and electrons constrained to bound orbits by the geomagnetic field in the manner visualized in classical theoretical studies by H Poincaré, C Stoermer, and H. Alfvén.

Radiation in the inner zone is quite different in composition from that in the outer zone. In the heart of the inner zone at an altitude of about 3600 km on the geomagnetic equator, the approximate composition includes electrons of energy greater than 40 kev having a maximum of unidirectional intensity of $\sim 2 \times 10^9/(\text{cm}^2)(\text{sec})(\text{steradian})$; electrons of energy greater than 600 kev which have a maximum of unidirectional intensity of $\sim 1 \times 10^7/(\text{cm}^2)(\text{sec})(\text{steradian})$, and protons of energy greater than 40 Mev which have an omnidirectional intensity of $\sim 2 \times 10^4/(\text{cm}^2)(\text{sec})$ The inner zone is observed to be relatively free of temporal fluctuations.

The outer zone, as mentioned earlier, undergoes marked temporal fluctuations It appears that *Pioneer* IV passed through the outer zone during a period of exceptional enhancement (March 3, 1959) The composition of the radiation in the heart of the outer zone (altitude about 20 000 km on the geomagnetic equator) on that date was approximately as follows (1) electrons of energy greater than 40 kev having an omnidirectional intensity of $\sim 1 \times 10^{11}/(\text{cm}^2)(\text{sec})$, (2) electrons of energy greater than 200 kev having an omnidirectional intensity less than $1 \times 10^8/(\text{cm}^2)(\text{sec})$, and (3) protons of energy greater than 60 Mev having an omnidirectional intensity less than $1 \times 10^2/(\text{cm}^2)(\text{sec})$ There was no significant information for protons of energy less than 30 Mev

Origin of trapped particles. It now appears likely that the protons and electrons in the inner zone are the products of the radioactive decay of neutrons moving outward from the top of the atmosphere These neutrons are released from cosmic-ray induced nuclear disintegrations in the atmosphere

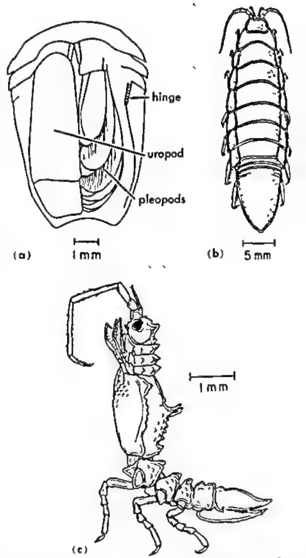
The outer zone almost certainly owes its existence to ionized solar gas ejected from the Sun This gas may travel to the Earth and be injected into the geomagnetic field Presumably, the outer zone contains large densities of protons and elec-

Timing. At low piston speeds gas friction and inertia are small and the valves should be opened or closed as the piston reaches the end of the appropriate stroke. At high piston speeds the inertia of the charge in the inlet passages causes the gas to continue to flow into the cylinder long after the piston has started up on the compression stroke. To trap the maximum amount of fresh charge in the cylinder it is therefore necessary to delay the closure of the inlet valve up to about 60° after bottom center, depending upon the geometry of the inlet passages and the piston speed. Late inlet closing increases the torque or pulling power of the engine at high speeds, but reduces torque at low speeds, because of the loss of fresh charge back into the inlet system during compression. To reduce the work required to expel the products of combustion during the exhaust stroke, it is necessary, at high speeds, to open the exhaust valve considerably before bottom center so that most of the gas is blown out and the cylinder pressure has fallen nearly to atmospheric before the exhaust stroke begins. In supercharged engines the inlet valve often opens near top center before the exhaust valve closes. The time during which both valves are open is called the overlap period. Overlap permits scavenging of the combustion space with fresh charge, for valve cooling and increased power, but results in poor idling characteristics because combustion products are drawn back through the engine by the low pressure of the inlet system when the throttle is nearly closed. [A R R.]

Valvifera

A suborder of isopod crustaceans resembling woodlice and pill bugs (*see* ISOPODA). They are distinguished from these and other isopods by having a *pair of flat, valvelike uropods which hinge laterally* and fold inward beneath the rear part of the body. The uropods open like doors to reveal five pairs of leaflike appendages, or pleopods, which are used for respiration and swimming. Members of the group are almost exclusively marine. Some species are omnivorous and have been found in dense populations on accumulations of broken algae and fish waste. Many thousands are sometimes taken in a single short dredge haul, and in these numbers they serve as effective scavengers in the economy of the sea. One species, *Idotea neglecta*, has been reported as a serious pest of catches of fish retained alive in keep nets in the harbor at Bergen, Norway.

The suborder is divided into two main families, the Idoteidae and the Arcturidae. Idoteidae have the body more or less flattened, as is typical in isopods. The seven pairs of walking legs are all similar, though those on the posterior segments are longer and more slender. Species in this family are more common inshore than in great depths and can readily be found on intertidal seaweeds or recently washed up wrack. Others may also be found in surface drift weed. Specimens of the American *Idotea metallica*, which is not a British form, are regularly found off the west coasts of Britain. They are presumably carried from the east coast of the United



(a) The under surface of the posterior end of *Idotea*, with one uropod removed at the hinge to reveal the pleopods (*Marine Biological Assoc.*). (b) *Idotea neglecta*. (c) *Arcturella*, a member of the family Arcturidae (*from G. O. Sars, An Account of the Crustacea of Norway, vol. 2, 1897*).

States on drifting weed and debris in the North Atlantic drift. Several forms in this family are able to tolerate brackish water, and one race of the species *Mesidotea entomon* is even established in fresh-water lakes in Sweden. This species is of particular interest in the study of the manner in which fresh-water races have evolved from brackish-water forms. Physiological and geological evidences show that the colonization of fresh water by *Mesidotea* has taken place since the last ice age, but the method by which osmoregulation occurs has not yet been determined.

Arcturidae have an almost cylindrical body and possess extremely long antennae. Only the last three pairs of legs compare with those of Idoteidae; the first four pairs are fragile and without claws, and they project forward beneath the head. Many species have the habit of clinging to an object with the posterior legs and moving the front part to and fro in search of food, which is seized by the antennae. The mobility of the ant[illegible] is increased in some species by the e[illegible]on

Operating voltage. Voltages achieved by generators of this type are proportional to the radius of the spherical high voltage terminal A well-constructed machine with a terminal of 1-m radius will operate up to about 10^6 volts. For good performance, the terminal should be distant from any object such as a wall, and the humidity of the air must be low.

A large proportion of the Van de Graaff Generators constructed since 1935 are enclosed by steel tanks containing gas such as air at a pressure of 10–20 atm The electrical gradient that gas will withstand without sparkover increases as gas pressure is increased Thus, by the use of high-pressure gas, the apparatus can be made smaller for a given voltage.

Although the modern machine operating in high pressure gas, as illustrated in Fig. 2, is the same in principle as the open-air device, the detailed requirements are more severe and the machine is

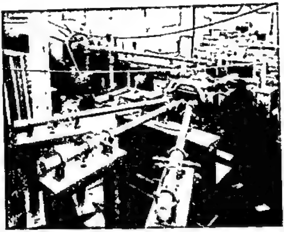


Fig. 3 Machine produced by High Voltage Engineering Corporation for acceleration of negative ions to give energy doubling. This machine gives proton beams with energies above 13 Mev. The machine proper is enclosed in the large tank in the upper left part of the photograph The proton beams go down the pipe attached to the tank, are magnetically deflected, and strike the target at the end of one of the three end-tubes (*High Voltage Engineering Corp*)

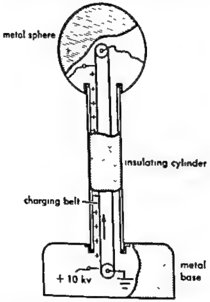


Fig 1 Schematic drawing of Van de Graaff Generator for operation in air at atmospheric pressure

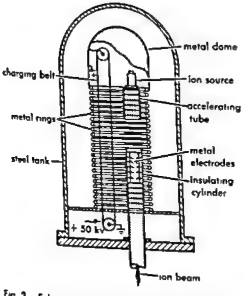


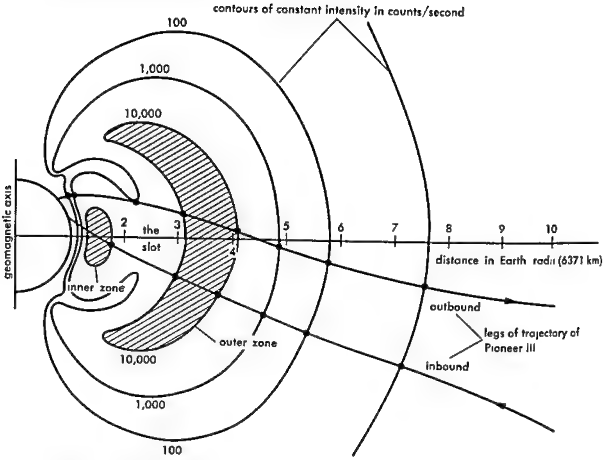
Fig 2 Schematic drawing of Van de Graaff Generator for operation in high-pressure gas

uniform potential drop between adjacent rings. Thus a uniform electric field is maintained along the insulating supports and the charging belt. Insulating supports are subdivided by metal conductors. Since a short insulator will withstand a higher gradient than a long insulator, electrical subdivision results in better insulator performance.

Most machines constructed after 1940 use gases other than air for the insulating medium because of the fire hazard Carbon dioxide or freon (CCl_2F_2) and nitrogen are commonly used The maximum voltage at which most Van de Graaff Generators will operate is set by electrical failure in the accelerating tube. These tubes are made up of rings of porcelain or glass, 1–2 in long, separated by metal disks

Van de Graaff Generators have been available commercially since about 1946 from the High Voltage Engineering Corporation Their positive ion accelerators range in size from a 1,000,000-volt machine housed in a tank 3 ft in diameter and 5 ft tall to a 5,500,000-volt unit enclosed by a tank 8 ft in diameter and 24 ft tall

Negative-ion accelerators. In 1958, this corporation completed a machine (Fig 3) for acceleration of negative ions Negative hydrogen ions generated in a special ion source are accelerated as they pass from ground potential up to the high-voltage termi-



Section through Earth's radiation belts.

trons whose energies lie below the detection limits of instruments used thus far The site and mechanism of acceleration of outer-zone particles to the energies observed are not yet known. Evidence, although not conclusive, favors local acceleration within the outer reaches of the fluctuating geomagnetic field It might conceivably be that the geomagnetic fluctuations induced by the arrival of solar plasma accelerate electrons and ions already present in the Earth's ionosphere.

Trapped particles eventually leak out of the radiation zones by virtue of collisional energy loss and scattering in the thin outer atmosphere and by magnetic scattering in the fluctuating geomagnetic field. Representative values of the mean trapping time for protons and electrons of various energies lie in the range of 10^5–10^9 sec. The observed intensities at any one time are thus the quasi-equilibrium values in a dynamic physical situation *See* COSMIC ELECTRODYNAMICS; GEOMAGNETISM; MAGNETOHYDRODYNAMICS; SOLAR RADIATION [J A.V.A]

Van de Graaff Generator

A high-voltage electrostatic generator in which electrical charge is carried from ground to a high-voltage terminal by means of an insulating belt. Current capacity is limited to a few milliamperes. Each generator is equipped with an evacuated tube through which charged particles of any type may be accelerated Machines equipped with positive ion sources and arranged for positive ion acceleration are widely used for research applications. Machines equipped with electron sources are used for x-ray therapy, industrial radiography, food and drug sterilization, and for research.

Operation. Principles of operation of the Van de Graaff Generator are most easily understood from study of a simple model as shown in Fig 1. The charging belt, which is made of insulating material such as rubberized fabric, is carried by two pulleys One, at ground potential, is driven by an electric motor, and the second is in the high-voltage terminal. The terminal is a metal enclosure which must be smooth and well rounded over its exterior surface It may be spherical in shape and an opening must be provided through which the charging belt passes The terminal is supported by a cylinder of insulating material. The base is enclosed by metal with a smooth, well-rounded exterior surface, and this enclosure is at ground potential Charge is sprayed onto the belt from a row of needle points or a fine wire held at a potential of about 10 kilovolts (kv). If the needle points are at a positive potential, they serve to charge the belt positively; if negative, they spray negative charge onto the belt. The pulley at high potential must be well within the enclosing electrode where the field due to the charge on the electrode is essentially zero A row of needles near the upper pulley with points close to the belt will remove charge from the belt, and the charge will pass to the outer surface of the terminal. Terminal voltage will rise to some limit determined by sparkover through air, by insulator failure, or by a load such as ion [illegible]ugh an evacuated tube.

Vanadium metal (calcium reduction). The best-known process for producing ductile vanadium metal now used in this country is the calcium reduction of the oxide, as developed by R. K. McKechnie and A. U. Seybolt. This process consists of charging pure vanadium oxide, calcium metal, and iodine into a heavy-walled steel cylinder, taking special precautions to exclude moisture. The cylinder is sealed and evacuated, and the reaction between the ingredients is initiated by the application of heat. Within a short time, high enough temperature and pressure are reached to allow the molten droplets of vanadium to collect beneath the calcium oxide-calcium iodide slag and to form a single button or regulus. A typical analysis of the product is 99.7% vanadium, 0.10% oxygen, 0.04% nitrogen, 0.008% hydrogen, 0.01% iron, and 0.03% carbon.

Vanadium metal (magnesium reduction). The process for producing pure vanadium by the magnesium reduction of vanadium trichloride is similar to the Kroll process ...

... duced with magnesium in an atmosphere of argon at a temperature of about 1550°F, resulting in a sponge which is subsequently separated from the magnesium chloride and excess magnesium by melting under argon, then by heating under vacuum, and finally by leaching. A typical analysis of the sponge is said to be 99.7% vanadium, 0.12% oxygen, 0.005% nitrogen, 0.01% hydrogen, 0.01% magnesium, and 0.03% iron.

Properties of ductile vanadium. In its pure form vanadium is soft and ductile. It can be hot- and cold-worked easily, but it must be heated in an inert atmosphere or in a vacuum because it oxidizes readily at temperatures above the melting point of its oxide, about 1225°F. The strength of vanadium is sensitive to interstitial impurities and varies from 30 000 psi in the purest form to 80,000 psi in the commercial form. This metal retains its strength unusually well at elevated temperatures. Vanadium has a relatively low cross section for neutron capture, 5 barns. This property has accounted for considerable interest in connection with the utilization of atomic energy.

Its resistance to hydrochloric and sulfuric acids is outstanding ...
attack bett...
cannot wi...
centrated.

Vanadium has a density of 6.1 g/cm³ or 0.22 lb/in.³, which is 22% less than that of iron and ...
more than ...
cal an...
higher ...
... elastic modulus of vanadium at room temperature is of the order of 20 000 000 psi.

Alloys. Almost all the vanadium produced is consumed by the steel industry as an alloying agent. The formation of vanadium carbide when vanadium is added to steel is the basis for many of the unique properties imparted to steel by vanadium. These carbides are extremely hard and wear-resistant; they do not coalesce readily but maintain a state of fine dispersion. A relatively small amount of vanadium (0.06–0.10%) is soluble in the austenitic phase of steels. This small percentage markedly increases the ability of the steel to harden on rapid cooling. Vanadium also forms a stable nitride, and it can, in effect, lower the nitrogen content of steel.

Tool steels are the largest class of vanadium-containing steels. In fact, almost all tool steels do contain this element, the amount ranging from 0.10 to 5.00%. Vanadium is required in these steels to ensure the retention of hardness and cutting ability at the elevated temperatures generated by rapid cutting of metals.

Many large steel forgings contain vanadium in the range of 0.05–0.15%, and here it acts as a grain refiner as well as improving the mechanical properties of the forgings. In large steel castings as well as forgings, a small percentage of vanadium is unique in its favorable action of raising strength and ductility.

Wrought constructional steels in the form of bars, plates, and tubing often contain small quantities of vanadium to raise strength, impact resistance, weldability, and resistance to softening at moderately elevated temperatures.

Vanadium is used in cast iron to control the size and distribution of graphite flakes and to improve strength and wear resistance. It has also been found to be an effective alloying element for titanium-base alloys as well as for aluminum base alloys. *See* FERROALLOY.

Principal compounds. Like other elements in the ...

... readily form condensed acids.

The most important compounds of vanadium are vanadium pentoxide, V_2O_5, and ammonium metavanadate, NH_4VO_3. The pentoxide is commonly sold as the sodium salt of hexavanadic acid, $Na_2H_2V_6O_{17}$. Ammonium metavanadate is formed by adding excess ammonium chloride to an alkaline solution of vanadium pentoxide. Pure vanadium pentoxide, 99.6% V_2O_5, is made by calcining ammonium metavanadate.

Many other compounds have been produced on a smaller scale for study. These include chlorides, sulfides, nitrates, acetates, oxalates, nitrides, carbides, bromides, iodides, and fluorides. Vanadyl linoleate, oleate, palmitate, phenolate, resinate, and stearate are some of the metalloorganic compounds that have been made.

nal. Here both electrons are stripped from the negative hydrogen ion, and the proton is again accelerated as it passes to ground potential. The machine illustrated in Fig 3 accelerates protons to energies above 13 Mev. *See* ELECTROSTATICS; PARTICLE ACCELERATOR. [R.G.H.]

Bibliography: J. D. Craggs and J. M. Meek, *High Voltage Laboratory Technique*, 1954; D. Halliday, *Introductory Nuclear Physics*, 2d ed, 1955; E. Segre (ed), *Experimental Nuclear Physics*, vol. 3, 1959; A. K. Solomon, *Why Smash Atoms?*, rev. ed., 1946.

Van der Waals equation

An empirical relationship useful for a gas at high pressure. The equation between pressure, volume, and temperature was deduced by the Dutch physicist J. D. van der Waals. For a homogeneous gas at pressure p and volume v

$$(p + a/v^2)(v - b) = R_u T$$

where R_u is the universal gas constant, T is absolute temperature, and a and b are constants, these constants being different for different gases. The term a/v^2 accounts for the cohesion between molecules, and b^2, the co-volume, accounts for the volume occupied by the molecules themselves

At large specific volume, $a/v^2 \ll p$ and $b \ll v$ and the equation reduces to the equation of state for an ideal gas $p/v = R_u T$. For a van der Waals gas, the full equation is a cubic in v having three roots The temperature at which the three roots are equal is the critical point of the gas However, measurements of pressure, volume, and temperature on a real substance at the critical point do not yield consistent values for a and b *See* STEAM; THERMODYNAMIC PRINCIPLES. [G A H.]

Vanadate

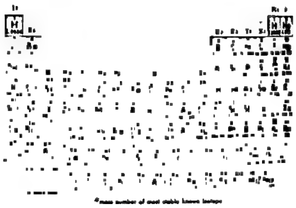
A generic term referring to salts containing vanadium in the 5+ oxidation state and derived from vanadium pentoxide, V_2O_5. Vanadium pentoxide is not very soluble in water, but salts corresponding to the acids, H_3VO_4, $H_4V_2O_7$, and HVO_3, which have been named as the corresponding phosphates (that is, ortho-, pyro-, and metavanadate) are known

These condensed ions are more ionic in character than the phosphates, which are described as polymers held together by covalent bonds

Mercury(I) metavanadate, $Hg_2(VO_3)_2$, for example, is a well-characterized salt that is only slightly soluble in water, and hence, is used in the quantitative analysis of vanadium. *See* PHOSPHATE; VANADIUM. [F.E.WR.]

Vanadium

Chemical element number 23, vanadium, V, is a metal which has been used primarily as an alloying element for steels and irons since about 1900 Several of its compounds are used in the chemical industry, notably in making oxidation catalysts, and in the ceramic industry as coloring agents. It was not until 1950 that this element could be produced in a sufficiently pure form and in large enough quantities to permit its study as an engineering material.



Natural occurrence. Although ores containing small amounts of vanadium are rather widely distributed throughout the world, the most important ones are found in the Western Hemisphere They include carnotite, roscoelite, vanadinite, and patronite Of these, carnotite is of the greatest commercial importance By far the largest tonnage of vanadium ore is mined in the United States.

Peak world vanadium output has been estimated at about 4500 metric tons in 1943, excluding that produced in the U.S.S.R. Consumption was largely in the form of ferrovanadium *See* CARNOTITE

Extraction from ores. The method most widely used to extract vanadium from carnotite ore is a roast-quench-leach process The ore is crushed to a size less than 10 mesh and roasted in a multiple-hearth furnace to form the water-soluble sodium metavanadate, $NaVO_3$. The hot ore is then quenched in water or soda ash solution. At this point, solvent extraction is employed to enrich the solution prior to the precipitation of sodium hexavanadate, $Na_2H_2V_6O_{17}$.

When dried, sodium hexavanadate is commonly known as red cake. It contains 85–90% V_2O_5, 2–10% Na_2O, and about 3% H_2O Solution of red cake with soda ash forms sodium vanadate, $NaVO_3$, which is filtered and precipitated with ammonia to form ammonium metavanadate, NH_4VO_3, a commonly used salt Heating of ammonium metavanadate drives off the ammonia and leaves a purified oxide, V_2O_5 For the production of alloys by smelting, red cake is fused and cast into flake form.

Ferrovanadium (reduction methods). Ferrovanadium is produced by carbon, aluminum, or silicon reduction of the oxide in the presence of iron in an electric-arc furnace Carbon reduction results in a product that contains relatively large amounts of both carbon and silicon, whereas silicon reduction requires a two-stage reduction to achieve an efficient operation Aluminum reduction of the oxide is commonly practiced; the reaction is exothermic and little additional power from the arc is required.

Vanadium metal (calcium reduction). The best-known process for producing ductile vanadium metal now used in this country is the calcium reduction of the oxide, as developed by R. K. McKechnie and A. U. Seybolt. This process consists of charging pure vanadium oxide, calcium metal, and iodine into a heavy-walled steel cylinder, taking special precautions to exclude moisture. The cylinder is sealed and evacuated, and the reaction between the ingredients is initiated by the application of heat. Within a short time, high enough ... reached to allow the

... typical analysis of the

carbon

Vanadium metal (magnesium reduction). The process for producing pure vanadium by the magnesium reduction of vanadium trichloride is similar to the Kroll process presently used to reduce titanium tetrachloride to titanium sponge. Vanadium trichloride is produced by the chlorination of ferrovanadium, or, as a later modification, by chlorination of vanadium pentoxide. The trichloride is reduced with magnesium in an atmosphere of argon at a temperature of about 1550°F, resulting in a sponge which is subsequently separated from the magnesium chloride and excess magnesium by melting under argon, then by heating under vacuum, and finally by leaching. A typical analysis of the sponge is said to be 99.7% vanadium, 0.12% oxygen, 0.005% nitrogen, 0.01% hydrogen, 0.01% magnesium, and 0.03% iron.

Properties of ductile vanadium. In its pure form vanadium is soft and ductile. It can be hot- and cold-worked easily, but it must be heated in an inert atmosphere or in a vacuum because it oxidizes readily at temperatures above the melting point of its oxide, about 1225°F. The strength of vanadium is sensitive to interstitial impurities and varies from 30,000 psi in the purest form to 80,000 psi in the commercial form. This metal retains its strength unusually well at elevated temperatures. Vanadium has a relatively low cross section for neutron capture, 5 barns. This property has accounted for considerable interest in connection with ...

... attack better than most stainless steels. Vanadium cannot withstand nitric acid, either dilute or concentrated.

Vanadium has a density of 6.1 g/cm³ or 0.22 lb/in.³, which is 22% less than that of iron and 28% more than ...

... vanadium at room temperature is of the order of 20,000,000 psi.

Alloys. Almost all the vanadium produced is consumed by the steel industry as an alloying agent. The formation of vanadium carbide when vanadium is added to steel is the basis for many of the unique properties imparted to steel by vanadium. These carbides are extremely hard and wear resistant; they do not coalesce readily but maintain a state of fine dispersion. A relatively small amount of vanadium (0.06–0.10%) is soluble in the austenitic phase of steels. This small percentage markedly increases the ability of the steel to harden on rapid cooling. Vanadium also forms a stable nitride, and it can, in effect, lower the nitrogen content of steel.

Tool steels are the largest class of vanadium-containing steels. In fact, almost all tool steels do contain this element, the amount ranging from 0.10 to 5.00%. Vanadium is required in these steels to ensure the retention of hardness and cutting ability at the elevated temperatures generated by rapid cutting of metals.

Many large steel forgings contain vanadium in the range of 0.05–0.15%, and here it acts as a grain refiner as well as improving the mechanical properties of the forgings. In large steel castings as well as forgings, a small percentage of vanadium is unique in its favorable action of raising strength and ductility.

Wrought constructional steels in the form of bars, plates, and tubing often contain small quantities of vanadium to raise strength, impact resistance, weldability, and resistance to softening at moderately elevated temperatures.

Vanadium is used in cast iron to control the size and distribution of graphite flakes and to improve strength and wear resistance. It has also been found to be an effective alloying element for titanium-base alloys as well as for aluminum-base alloys. See FERROALLOY.

Principal compounds. Like other elements in the transition group Va, vanadium forms numerous and frequently complicated compounds because of its variable valence. It has at least four oxidation states, 2+, 3+, 4+, and 5+. It is amphoteric, although predominantly basic in the lower oxidation states, acidic in the higher ones. It forms derivatives from more or less well-defined radicals such as VO^{2+} and VO^{3+}. Its oxygen acids readily form condensed acids.

The most important compounds of vanadium are vanadium pentoxide, V_2O_5, and ammonium metavanadate, NH_4VO_3. The pentoxide is commonly sold as the sodium salt of hexavanadic acid, $Na_2H_2V_6O_{17}$. Ammonium metavanadate is formed by adding excess ammonium chloride to an alkaline solution of vanadium pentoxide. Pure vanadium pentoxide, 99.6% V_2O_5, is made by calcining ammonium metavanadate.

Many other compounds have been produced on a smaller scale for study. These include chlorides, sulfides, nitrates, acetates, oxalates, nitrides, carbides, bromides, iodides, and fluorides. Vanadyl linoleate, oleate, palmitate, phenolate, resinate, and stearate are some of the metalloorganic compounds that have been made.

The largest application of vanadium compounds is in the manufacture of oxidation catalysts for the chemical industry. Both ammonium metavanadate and vanadium pentoxide are used for this purpose. Processes which employ such catalysts include the manufacture of polyamides, such as nylon, the manufacture of sulfuric acid by the contact process, the manufacture of phthalic and maleic anhydrides, and various other organic oxidation reactions such as anthracene to anthraquinone, alcohol to acetaldehyde, diphenylamine to carbazole, and sugar to oxalic acid.

Vanadium in the pentavalent state is used in clear glass to absorb ultraviolet rays; the passage of wavelengths below 3589 angstroms is prevented. The addition of only 0.02% V largely removes the harmful actinic rays which injure the eye and fade fabrics. When it is present in larger amounts, vanadium pentoxide produces a greenish-yellow color in glass.

Vanadium compounds are likewise used in the ceramic industry for glazes and enamels. Zirconia with silica and vanadium pentoxide produces blue to bluish-green color. Vanadium yellow is produced by heating lead oxide with zirconia and vanadium salts. A yellow fluorescent product is made from zinc oxide and vanadium pentoxide. Tin oxide and vanadium oxide produce a stable yellow color which can be used for glaze, underglaze, or body stain. Cadmium sulfide with selenium and vanadium compounds yields a bright red color.

In dye manufacture and dyeing, vanadium compounds are widely used in the production of aniline black. Here, vanadium salts are added as catalysts to a mixture of aniline hydrochloride and potassium or sodium chlorate. A very rich black is produced and the over-all cost is low. Vanadium compounds are also used as mordants for the dyeing and printing of cotton, and particularly for fixing aniline black on silk. Certain inks are given a lasting quality by a small addition of ammonium metavanadate. See METALLOACID ELEMENTS, TRANSITION ELEMENTS. [T.W.M.]

Bibliography: C. A. Hampel (ed.), *Rare Metals Handbook*, 1954.

Vanilla

A choice flavoring obtained from a climbing orchid, *Vanilla fragrans*, a native of tropical American forests. Its fruits are pods called vanilla beans. These are picked at the proper time before they have fully matured and then subjected to a prolonged and critical curing process. During the curing, enzymatic action converts a glucoside into vanillin which has the characteristic odor and flavor. Vanilla is used in cookery, confectionery, and beverages. Vanilla extract, most used, is prepared by extracting the crushed beans with alcohol. A synthetic vanillin is made from eugenol occurring in clove oil, but the natural product is preferred. Several plants have been used as substitutes for true vanilla but these are of little value. See ORCHIDALES; SPICE AND FLAVORING. [P.D.S.]

Vanilla (Vanilla fragrans). (USDA)

Vanilla extract

An ethyl alcohol extract of the vanilla bean (*Vanilla planifolia*) used in the food industry for flavoring foods. The pod, 6–8 in. long, is the fruit of a member of the orchid family. See SPICE AND FLAVORING; *see also* FOOD ENGINEERING.

Vanilla is native to Central America, but has been cultivated in Madagascar, the source of 80% of United States imports. Other sources include Reunion and Comores Islands as well as Mexico, Tahiti, Java, and South America.

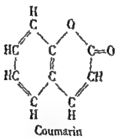
Vanilla beans are harvested as a full grown, unripe fruit from the vines which bear them. The beans are placed in piles and allowed to ferment. Heat developed during fermentation produces optimum conditions for flavor development. After alternate sweating and drying, which occupies several weeks and during which the beans turn dark brown, the beans are sun dried, coated with oil and packed in bundles. The surface of the beans frequently turns chalky white with a surface precipitation of vanillin, the principal flavoring constituent.

Vanilla extract is prepared from beans which have been cut into small pieces, then bruised. The cut beans are percolated with alcohol, to obtain the extract.

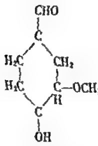
Vanilla flavor is due principally to vanillin, but there are other important constituents. A simple extract of vanillin in alcohol with caramel coloring closely resembles true vanilla. But this lacks the minor constituents and is readily detectable from the true bean extract. Extracts from the tonka bean (*Dipteryx odorata*), similar to vanilla extracts, are high in coumarin rather than vanillin.

Vanillin is readily synthesized and is manufactured in large quantities for use in artificial vanilla extract. It was once made [illegible] al altera-

tion of eugenol, but is now largely made from the wood product lignin True vanilla extract must contain the extract from 10 g of beans in 100 ml of extract Vanilla may also contain sugar, glycerol, or dextrose



Coumarin



Vanillin

Vanilla extract is used extensively by the chocolate and confectionery industry and as a flavor in ice cream, baked goods, and soft drinks. It is bottled and sold for use in home cooking *See* COCOA POWDER AND CHOCOLATE. [R F M]

Bibliography: M B Jacobs (ed), *Chemistry and Technology of Food and Food Products*, 2 vols, 1944, L. W Jones (ed), *A Treasury of Spices*, 1956

Vapor cycle

A thermodynamic cycle, operating as a heat engine or a heat pump, during which the working substance is in, or passes through, the vapor state A vapor is a substance at or near its condensation point. It may be wet, dry, or slightly superheated One hundred per cent dryness is an exactly definable condition which is only transiently encountered in practice Vapor behavior deviates so widely from the ideal gas laws that calculation requires the use of tables and ... mentally de...

Power at ... plant opera... on a vapor cycle where steam is generated by boiling water at high pressure, expanding it in a prime mover, exhausting it to a condenser where it is reduced to the liquid state at low pressure, and then returning the water by a pump to the boiler (Fig. 1)

In the customary vapor compression refrigeration plant, the process is essentially reversed with the refrigerant evaporating at low temperature and pressure, being compressed to high pressure, condensed at elevated temperature and returned as liquid refrigerant through an expansion valve to the evaporating coil (Fig 2).

The Carnot cycle, between any two temperatures, gives the limit for the efficiency of the conversion of heat into work (Fig. 3). This efficiency is independent of the properties of the working fluid. Although the thermal efficiency is independent of the properties of the fluid, the mean effective pressure (and consequent physical dimensions of the engine) will be vitally influenced by choice of fluid (compare Fig 3*b* with Fig 3*c*). The Carnot cycle is not realistic for the evaluation of steam power plant performance because the cycle precludes the use of superheat and calls for the isentropic compression of vapor It is useful, however, for specifying the limiting efficiency that a real cycle might approach It is so used in judging performance of vapor-cycle heat engines and heat pumps.

The Rankine cycle is more realistic in describing the ideal performance of steam power plants and vapor compression refrigeration systems.

Vapor steam plant. In the case of the steam power plant (Fig. 1), the Rankine cycle (Fig 4) has two constant pressure phases joined by a reversible adiabatic (isentropic) phase (1-2) From the properties of the fluid, the work of the prime mover, ΔW_{PM}, is most conveniently evaluated as

$$\Delta W_{PM} = h_1 - h_2$$

where h is the enthalpy, Btu/lb The feed pump uses some of this work, ΔW_{FP}, to return the water from the condenser to the boiler so that the net output of the cycle is $\Delta W_{PM} - \Delta W_{FP}$ This net output can be related to the heat that must be

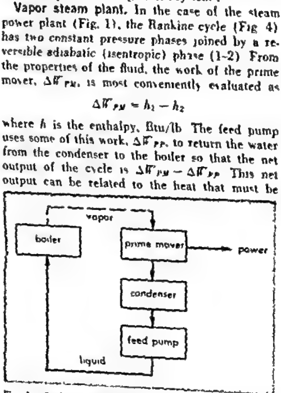


Fig. 1 Rudimentary steam power plant flow diagram.

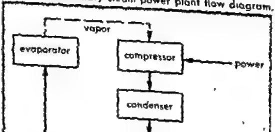


Fig 2. Rudimentary vapor-compression refrigeration plant flow diagram

The largest application of vanadium compounds is in the manufacture of oxidation catalysts for the chemical industry. Both ammonium metavanadate and vanadium pentoxide are used for this purpose. Processes which employ such catalysts include the manufacture of polyamides, such as nylon, the manufacture of sulfuric acid by the contact process, the manufacture of phthalic and maleic anhydrides, and various other organic oxidation reactions such as anthracene to anthraquinone, alcohol to acetaldehyde, diphenylamine to carbazole, and sugar to oxalic acid.

Vanadium in the pentavalent state is used in clear glass to absorb ultraviolet rays; the passage of wavelengths below 3589 angstroms is prevented. The addition of only 0.02% V largely removes the harmful actinic rays which injure the eye and fade fabrics. When it is present in larger amounts, vanadium pentoxide produces a greenish-yellow color in glass.

Vanadium compounds are likewise used in the ceramic industry for glazes and enamels. Zirconia with silica and vanadium pentoxide produces blue to bluish-green color. Vanadium yellow is produced by heating lead oxide with zirconia and vanadium salts. A yellow fluorescent product is made from zinc oxide and vanadium pentoxide. Tin oxide and vanadium oxide produce a stable yellow color which can be used for glaze, underglaze, or body stain. Cadmium sulfide with selenium and vanadium compounds yields a bright red color.

In dye manufacture and dyeing, vanadium compounds are widely used in the production of aniline black. Here, vanadium salts are added as catalysts to a mixture of aniline hydrochloride and potassium or sodium chlorate. A very rich black is produced and the over-all cost is low. Vanadium compounds are also used as mordants for the dyeing and printing of cotton, and particularly for fixing aniline black on silk. Certain inks are given a lasting quality by a small addition of ammonium metavanadate. *See* METALLOACID ELEMENTS, TRANSITION ELEMENTS [T W M]

Bibliography: C. A. Hampel (ed.), *Rare Metals Handbook*, 1954.

Vanilla

A choice flavoring obtained from a climbing orchid, *Vanilla fragrans*, a native of tropical American forests. Its fruits are pods called vanilla beans. These are picked at the proper time before they have fully matured and then subjected to a prolonged and critical curing process. During the curing, enzymatic action converts a glucoside into vanillin which has the characteristic odor and flavor. Vanilla is used in cookery, confectionery, and beverages. Vanilla extract, most used, is prepared by extracting the crushed beans with alcohol. A synthetic vanillin is made from eugenol occurring in clove oil, but the natural product is preferred. Several plants have been used as substitutes for true vanilla but these are of little value. *See* ORCHIDALES; SPICE AND FLAVORING [P D S.]

Vanilla (*Vanilla fragrans*). (USDA)

Vanilla extract

An ethyl alcohol extract of the vanilla bean (*Vanilla planifolia*) used in the food industry for flavoring foods. The pod, 6–8 in. long, is the fruit of a member of the orchid family. *See* SPICE AND FLAVORING, *see also* FOOD ENGINEERING.

Vanilla is native to Central America, but has been cultivated in Madagascar, the source of 80% of United States imports. Other sources include Reunion and Comores Islands as well as Mexico, Tahiti, Java, and South America.

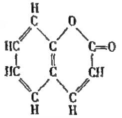
Vanilla beans are harvested as a full grown, unripe fruit from the vines which bear them. The beans are placed in piles and allowed to ferment. Heat developed during fermentation produces optimum conditions for flavor development. After alternate sweating and drying, which occupies several weeks and during which the beans turn dark brown, the beans are sun dried, coated with oil and packed in bundles. The surface of the beans frequently turns chalky white with a surface precipitation of vanillin, the principal flavoring constituent.

Vanilla extract is prepared from beans which have been cut into small pieces, then bruised. The cut beans are percolated with alcohol, to obtain the extract.

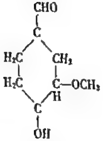
Vanilla flavor is due principally to vanillin, but there are other important constituents. A simple extract of vanillin in alcohol with caramel coloring closely resembles true vanilla. But this lacks the minor constituents and is readily detectable from the true bean extract. Extracts from the tonka bean (*Dipteryx odorata*), similar to vanilla extracts, are high in coumarin rather than vanillin.

Vanillin is readily synthesized and is manufactured in large quantities for use in artificial vanilla extract. It was once made ... ical altera

tion of eugenol, but is now largely made from the wood product lignin True vanilla extract must contain the extract from 10 g of beans in 100 ml of extract. Vanilla may also contain sugar, glycerol, or dextrose



Coumarin



Vanillin

Vanilla extract is used extensively by the chocolate and confectionery industry and as a flavor in ice cream, baked goods, and soft drinks It is bottled and sold for use in home cooking *See* COCOA POWDER AND CHOCOLATE [R.E M]

Bibliography: M B Jacobs (ed), *Chemistry and Technology of Food and Food Products*, 2 vols., 1941; L. W Jones (ed), *A Treasury of Spices*, 1956.

Vapor cycle

A thermodynamic cycle, operating as a heat engine or a heat pump, during which the working substance is in, or passes through, the vapor state A vapor is a substance at or near its condensation point. It may be wet, dry, or slightly superheated One hundred per cent dryness is an exactly definable condition which is only transiently encountered in practice. Vapor behavior deviates so widely from the ideal gas laws that calculation requires the use of tables and graphs that give the experimentally determined properties of the fluid

Power and refrigeration plants. A steam power plant operates on a vapor cycle where steam is generated by boiling water at high pressure, expanding it in a prime mover, exhausting it to a condenser where it is reduced to the liquid state at low pressure, and then returning the water by a pump to the boiler (Fig 1)

In the customary vapor-compression refrigeration plant the process is essentially reversed with the refrigerant evaporating at low temperature and pressure, being compressed to high pressure, condensed at elevated temperature, and returned as liquid refrigerant through an expansion valve to the evaporating coil (Fig 2)

The Carnot cycle, between any two temperatures, gives the limit for the efficiency of the conversion of heat into work (Fig. 3). This efficiency is independent of the properties of the working fluid Although the thermal efficiency is independent of the properties of the fluid, the mean effective pressure (and consequent physical dimensions of the engine) will be vitally influenced by choice of fluid (compare Fig. 3*b* with Fig 3*c*). The Carnot cycle is not realistic for the evaluation of steam power plant performance because the cycle precludes the use of superheat and calls for the isentropic compression of vapor. It is useful, however, for specifying the limiting efficiency that a real cycle might approach It is so used in judging performance of vapor-cycle heat engines and heat pumps.

The Rankine cycle is more realistic in describing the ideal performance of steam power plants and vapor compression refrigeration systems.

Vapor steam plant. In the case of the steam power plant (Fig. 1), the Rankine cycle (Fig. 4) has two constant pressure phases joined by a reversible adiabatic (isentropic) phase (1–2). From the properties of the fluid, the work of the prime mover, ΔW_{PM}, is most conveniently evaluated as

$$\Delta W_{PM} = h_1 - h_2$$

where h is the enthalpy, Btu/lb The feed pump uses some of this work, ΔW_{FP}, to return the water from the condenser to the boiler so that the net output of the cycle is $\Delta W_{PM} - \Delta W_{FP}$ This net output can be related to the heat that must be

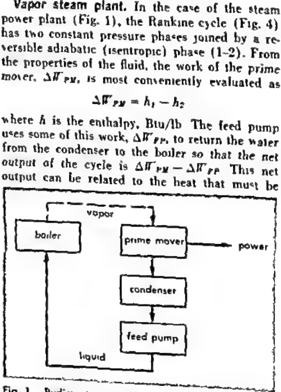


Fig 1. Rudimentary steam power plant flow diagram.

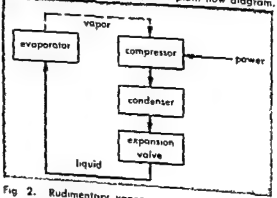


Fig 2. Rudimentary vapor-compression refrigeration plant flow diagram

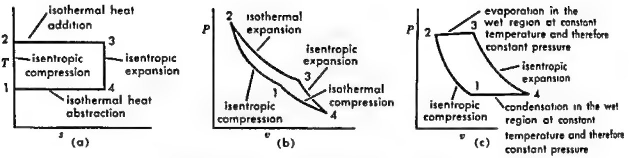


Fig 3 Carnot cycle (a) Temperature-entropy. (b) Pressure-volume for fixed gas. (c) Pressure-volume for vapor

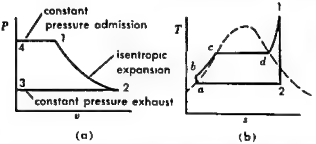


Fig. 4. Rankine cycle for heat-engine plant. (a) Pressure-volume diagram of prime mover (b) Temperature-entropy diagram for the power plant

added to produce steam by consideration of the T-s diagram (Fig. 4b). The area under line bcd-1 is the heat supplied in the boiler (heat source); phase 1–2 is the isentropic expansion in the prime mover; the area under line 2-a is the heat rejected to the condenser (heat sink); phase a-b is the isentropic compression of the liquid in the feed pump. Thus, the thermal efficiency is

$$\frac{\text{work done}}{\text{heat added}} = \frac{\Delta W'_{PM} - \Delta W'_{FP}}{h_1 - h_a - \Delta W'_{FP}}$$

There are many variables which influence the performance of the Rankine cycle. For steam the thermal efficiency is a function of pressure, temperature, and vacuum (Fig. 5). High pressure, high superheat, and high vacuum lead to high efficiency

Vapor refrigeration plant. The Rankine cycle can be used to evaluate the performance of vapor-compression system of refrigeration (Fig 2) A counterclockwise path is followed on the P-v and T-s cycle diagrams (Fig 6). The refrigerant enters the compressor as low-temperature, low-pressure vapor, (4-1); isentropic compression follows (1-2) and then high-pressure delivery (2-3) The work to drive the compressor is $h_2 - h_1$. The coefficient of performance cp which is essentially the reciprocal of thermal efficiency, is

$$cp = \frac{\text{refrigeration}}{\text{work done}} = \frac{h_1 - h_d}{h_2 - h_1}$$

for a cooling machine, and

$$cp = \frac{\text{heat delivered}}{\text{work done}} = \frac{h_2 - h_c}{h_2 - h_1}$$

for a warming machine.

The difference $h_1 - h_d$ is heat removed in the refrigerating coils, the area under phase (d-1) on the T-s diagram The difference $h_2 - h_c$ is the heat delivered to the condensing coils, the area under phase (2abc) Because the flow is throttled through the expansion valve, the enthalpy is constant, $h_c = h_d$ Some ideal performance values of a vapor-compression refrigeration system are plotted in Fig. 7 For further details on heat-pump vapor cycles, *see* HEAT PUMP. The remainder of this article is concerned with the vapor cycle as it is applied to power-generation purposes

Regenerative heat cycle. The regenerative cycle is a modification of the simple Rankine cycle. Feed water is heated by extracted steam (Fig. 8). As a result, less heat must be added in the boiler to evaporate a pound of steam and in turn, to deliver a kwhr of work output from the associated steam engine.

A cycle with a single stage of regenerative heating can be viewed as two Rankine cycles superimposed on one another (Fig 8b). In one the exhaust pressure and consequent temperature are substantially higher than in the other (point f versus point 2) The heat of condensation represented by the area under c-f can be used to raise the feed temperature from T_b to T_c The area under phase

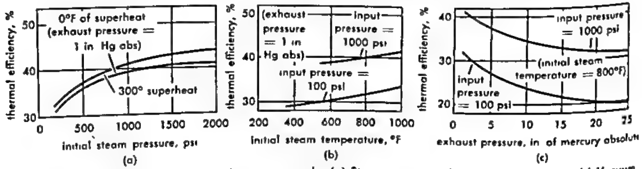


Fig 5. Thermal efficiency of ideal Rankine steam cycle. (a) Steam pressure. (b) Steam temperature. (c) Vacuum

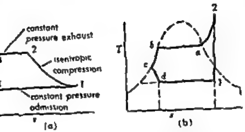


Fig 6 Rankine cycle for heat-pump plant (a) Pressure-volume diagram of compressor (b) Temperature-entropy diagram for a vapor-compression refrigeration plant

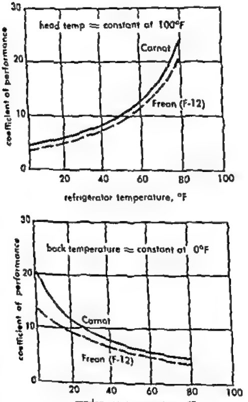


Fig 7 Coefficient of performance of ideal Rankine and Carnot cycle as influenced by condensing and refrigerating temperatures.

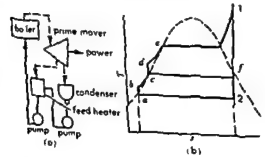


Fig 8 Regenerative cycle (a) Rudimentary steam power plant flow diagram with single stage feed heating (b) Temperature-entropy diagram

b-c is smaller than under phase *c-f* so a fraction of a pound of steam is needed to raise one pound of water to the common temperature level T_c.

The principle of regeneration can be extended to multiple stage heating with different final feed temperatures and in the limit reaching the boiler saturation temperature T_e with an infinite number of heating stages Some consequences of the process are reflected in the data of Fig. 9 The gain in thermal efficiency is the consequence of reducing the quantity of heat rejected to sink 2-*a* in Fig. 8. The weight of steam flow to the prime mover for the production of a kilowatt hour is larger than with the simple Rankine cycle (Fig 4) But the heat required to make a pound of steam is so much less that there is an over-all thermodynamic gain. Modern steampower practice uses the steam turbine for up to ten stages of regenerative heating.

Reheat cycle. The resuperheat or reheat cycle is another improvement in vapor cycles favored in current central station practice. Steam expanding isentropically (Fig 4*b*) grows wetter with consequent increased erosion of machinery parts and loss in mechanism efficiency. Superheat tends to correct these weaknesses but metallurgical limitations fix the maximum allowable steam temperature Reheating the steam after a partial expansion (Fig. 10*a*) gives a practical correction The reheating can be carried out at various pressure and temperature levels and in multiple stages Thermal efficiency is improved over the simple Rankine cycle (Fig 10*b*) Current practice uses a single reheat stage with temperatures approximately equal to the primary steam temperature Supercritical pressure plants favor two stages of reheating.

Binary vapor cycle. Comparison of the Rankine (Fig 4*b*), the reheat (Fig 10*a*), and Carnot (Fig. 3*a*) cycles shows that there are considerable thermodynamic losses in the first two by failure to approach the rectangular *T-s* configuration of the Carnot cycle The binary vapor cycle uses two fluids with totally different vapor pressures, such as mercury and water (Fig 11) If a Rankine cycle using

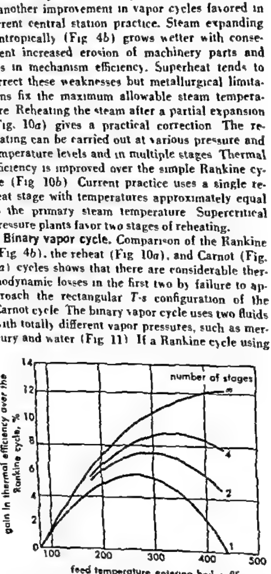
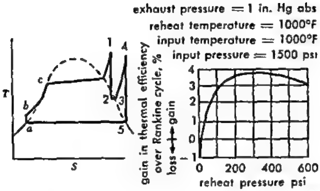


Fig 9 Gain in thermal efficiency by use of regenerative instead of Rankine cycle, steam conditions, 400 psi and 700°F, exhaust pressure 1 in Hg abs.



(a) (b)

Fig 10. Resuperheat or reheat cycle (a) Temperature-entropy diagram (b) Gain in thermal efficiency as function of reheat pressure; primary pressure, primary temperature, and reheat temperature constant at 1500 psi, 1000°F, and 1000°F, respectively.

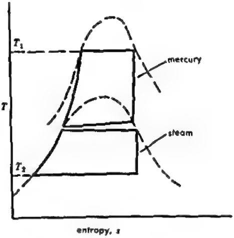


Fig 11. Binary-vapor cycle, temperature-entropy diagram

mercury is superimposed on that using steam. it is possible to operate the mercury-condenser as the steam-boiler, transferring the heat from the one fluid to the other Because of the differences in the latent heats and specific heats of the two fluids, several pounds of mercury must be circulated to make one pound of steam However, the combined *T-s* diagram (Fig 11) approximates the rectangular specification with appreciable gain in thermal efficiency. For many years the binary vapor cycle was thought to have practical as well as theoretical advantages, but its economic difficulties have led to its eclipse by, and abandonment in favor of, the high-pressure, regenerative, reheat, steam cycles [T B.]

Bibliography: T. Baumeister, *Marks' Mechanical Engineers' Handbook*, 6th ed, 1958, V. M Faires, *Thermodynamics*, 3d ed., 1957; G. A. Hawkins, *Thermodynamics*, 2d ed, 1951; J. H. Keenan, *Thermodynamics*, 1941.

Vapor lamp

An electrically operated source of radiant energy characterized by the emission of radiation from a stream of ionized gas carrying current between electrodes in the lamp. Vapor lamps (also called gaseous-discharge lamps) in common use in the United States include fluorescent, mercury vapor and neon lamps; in Europe, sodium vapor and xenon lamps also find broad use. In general, vapor lamps provide their characteristic energy at higher efficiencies than other sources; hence, they find wide use in lighting applications and where concentrations of ultraviolet energy are required *See* ELECTRICAL CONDUCTION IN GASES

The various types of vapor lamps all possess a negative resistance characteristic; that is, the resistance within the lamp envelope decreases with an increase in current. Without some form of current-limiting device in the electric circuit, current would rise swiftly after the lamp started until lamp failure occurred. This current-limiting element, or ballast, is external to the gas-discharge envelope, but is sometimes part of the complete lamp assembly. Resistance ballasting is used on direct-current circuits and on some alternating-current circuits. With most vapor lamps, resistance ballasting on ac circuits is inefficient because of the high proportion of circuit wattage dissipated in the ballast Ballasts using inductive or capacitive reactance are more efficient and more common. When lamps are operated on ac circuits with sufficient line voltage to start the lamps, a simple choke can be used as a ballast, when higher voltage is required, the ballast usually consists of an autotransformer with capacitors Some commercial and industrial fluorescent lighting systems operate at several hundred cycles per second, making capacitor ballasts advantageous in some applications.

For a more complete discussion of the characteristics of resistance, inductance, and capacitance elements in electrical circuits, *see* ALTERNATING CURRENT CIRCUIT THEORY; DIRECT-CURRENT CIRCUIT THEORY

For discussion of the common types of vapor lamps, *see* FLUORESCENT LAMP; MERCURY-VAPOR LAMP; NEON GLOW LAMP; SODIUM-VAPOR LAMP, *see also* ILLUMINATION [A M]

Vapor lock

Interruption of fuel flow to an engine due to blockage of passages in the fuel system by fuel vapor

To promote easy starting, all gasolines contain volatile constituents which under some conditions, such as high ambient temperature, tend to produce more vapor than the fuel pump and carburetor vents can handle The very action of the fuel pump in decreasing the pressure at its inlet, tends to vaporize the fuel If the vapor forms faster than the pump can draw it from the fuel line, the flow of fuel to the carburetor is effectively stopped and the engine stalls

The tendency to form vapor in a given fuel system has been correlated with the Reid vapor pressure and A S T M distillation curve of the gasoline (*see* AIRCRAFT FUEL) With a given fuel, vapor formation can be minimized by keeping all parts of the fuel system cool, eliminating sudden changes in cross section or direction of fuel lines and using a fuel pump of adequate [illegible]ty, at the lowest point in the fuel line (*see* [illegible]). [A R R]

Vapor pressure

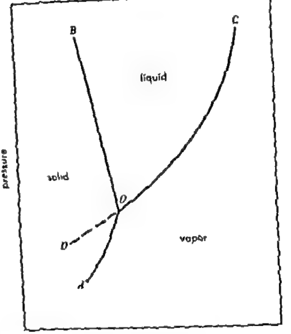
The saturation pressures exerted by vapors which are in equilibrium with their liquid or solid forms. One of the most important physical properties of a liquid, the vapor pressure, enters into many thermodynamic calculations, and underlies several methods for the determination of the molecular weights of substances dissolved in liquids. See MOLECULAR WEIGHT; SOLUTION. For a discussion of the vapor pressure relationships of solids, see SUBLIMATION.

If a liquid is introduced into an evacuated vessel at a given temperature, some of the liquid will vaporize, and the pressure of the vapor will attain a maximum value which is termed the vapor pressure of the liquid at that temperature. Although the quantity of liquid remaining does not diminish thereafter, the process of evaporation does not cease. A dynamic equilibrium is established, in which molecules escape from the liquid phase and return from the vapor phase at equal rates (see EVAPORATION).

[illegible]

pure liquid is a unique and characteristic property of the liquid and depends only upon the temperature. A gas or vapor may, on the other hand, exert any pressure [illegible]

[illegible] between the vapor pressure of a liquid and the temperature is indicated by a phase diagram. The illustration shows the phase diagram for water, the line OC being the vapor pressure line for liquid water. Line AO is the vapor pressure line (sublimation pressure curve) for ice, and line BO is the liquid-solid equilibrium line. Point O is called the triple point and is the unique pressure and temperature at which a pure solid, its liquid, and its vapor can coexist in equilibrium under the pressure of the vapor alone. The triple point of water is not the familiar melting point (0°C), but rather 0.0075°C. The distinction is that the total pressure at the triple point is 4.58 mm, the common vapor pressure of solid and liquid water, whereas the total pressure at the melting point is ordinarily 1 atm. Point C is called the critical point and is the point above which there [illegible] gaseous [illegible] at tempe[illegible] gardless [illegible] necessarily, the surface tension and the latent heat of vaporization become zero at the critical point. For ordinary liquids, the vapor pressure at the critical point is usually about 50 atm. There is no evidence for a critical point on the solid-liquid equilibrium line, and B is meant to indicate a direction, rather than a point; other phases may appear above B, of course, and alter the direction of the line. It is possible to undercool a liquid below its triple point if crystallization nuclei are absent. The vapor pressure of the undercooled liquid is given by the dashed line DO and lies above the equilibrium vapor pressure of the solid. The undercooled liquid is therefore metastable, because the system tends to assume its lowest vapor pressure at equilibrium.



Phase diagram of the water system.

Quantitative relations. For most liquids, the relationship between the vapor pressure and temperature can be expressed by an equation having the form

$$\log P = a + b \log T + \frac{c}{T}$$

where a, b, and c are constants. The simpler equation

$$\log P = \frac{m}{T} + n$$

where m and n are constants is often adequate. The change in the vapor pressure of a liquid with temperature may be expressed by the Clausius-Clapeyron equation

$$\frac{dP}{dT} = \frac{L_v}{T\Delta V}$$

where L_v is the molar latent heat of vaporization, and ΔV is the difference in molar volumes of the vapor and liquid at the temperature T. Because the molar volume of the liquid is negligible by comparison with that of the vapor except near the critical point, the Clausius-Clapeyron equation may be written

$$\frac{dP}{dT} = \frac{L_v}{TV}$$

where V is the molar volume of the gas. If the gas

follows the ideal gas law ($PV = RT$), the equation may be further written as

$$\frac{1}{P} \cdot \frac{dP}{dT} = \frac{L_v}{RT^2}$$

where R is the gas constant. For moderate ranges of temperature, the latent heat of vaporization is constant, and the equation may be rearranged and integrated to give the useful form

$$\log_{10} \frac{P_2}{P_1} = \frac{L_v}{2.3R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)$$

From the latent heat of vaporization and the vapor pressure at one temperature, the vapor pressures at other temperatures may thus be calculated.

The vapor pressure of a liquid is lowered when a substance is dissolved in it. At low solute concentrations in nonideal solutions, or at all solute concentrations in ideal solutions, the partial vapor pressure of the liquid in a solution is proportional to the mole fraction of the solvent:

$$P_1 = X_1 P_1^0$$

where X_1 is the solvent mole fraction, and P_1^0 is the vapor pressure of the pure solvent. Raoult's law, as this important relationship is known, may also be expressed in the form

$$\frac{P_1^0 - P_1}{P_1^0} = X_2$$

where x_2 is the mole fraction of the solute. The lowering of the vapor pressure of the solvent is proportional to the mole fraction of the solute. This relation provides the basis for the determination of the molecular weights of dissolved substances. See CONCENTRATION SCALES.

Vapor pressure measurement. Because the vapor pressures of liquids range widely, a number of methods have been devised for their measurement. The static method consists of introducing an excess of the liquid into an evacuated system at a given temperature, and measuring the vapor pressure with an attached mercury manometer. It is useful for vapor pressures ranging from a few millimeters of mercury up to several atmospheres. The dynamic method consists of introducing the liquid into a system in which the pressure may be varied, and noting the temperature at which the liquid boils at a given pressure. It is useful for moderately high vapor pressures. The transpiration method consists of bubbling a known volume of an inert gas at a definite pressure through the liquid. The quantity of liquid transported into the carrier gas is determined, and the vapor pressure of the liquid is calculated from the relation

$$P_L = P_C \frac{n_L}{n_L + n_C}$$

where P_C is the carrier gas pressure, and n_L and n_C are the numbers of moles of liquid vaporized and carrier gas, respectively. For the measurement of very low vapor pressures, the Knudsen effusion method may be used. This method is based upon the measurement of the mass of vapor which escapes through a very small hole in a vessel containing a liquid in equilibrium with its vapor. If the diameter of the hole is small compared with the mean free path of the gas molecules, the latter suffer no collisions in their passage through the hole. The following equation then relates the mass of gas which passes per cm²/sec through the hole and the vapor pressure,

$$m = P\sqrt{\frac{M}{2\pi RT}}$$

where m is the mass of gas per cm²/sec, and M is the molecular weight. The Knudsen method has been used with radioactive isotopes or with mass spectrometers to determine very low vapor pressures. See EQUILIBRIUM, PHASE; TRIPLE POINT.

[N.H.N.]

Bibliography: S. Glasstone, *Textbook of Physical Chemistry*, 2d ed., 1946; W. J. Moore, *Physical Chemistry*, 2d ed., 1955.

Varactor

A semiconductor diode designed to maximize the variation of its capacitance with applied voltage. Such diodes are often simply called variable-capacitance diodes and can be of either the junction or the point-contact variety. The variable capacitance arises from the fact that increasing the voltage drop across any semiconductor barrier causes a widening of the charge-depletion region of the barrier (see JUNCTION DIODE; SEMICONDUCTOR). This type of diode is used primarily in parametric amplifiers and subharmonic generators. [L.P.HU.]

Variable (discrete and continuous)

A term that became part of mathematical language during the development of analytic geometry (no equivalent word is found in the writings of Archimedes and other Greek mathematicians), and which is frequently not defined in modern works, apparently in the belief that the meaning ordinarily attached to the word is sufficiently precise. When a definition is given, it is usually to the effect that a symbol x is a variable if it may denote any member of a set S of objects. A variable is discrete or continuous according as its range (the set S) is discrete (for example, a subset of the natural numbers) or continuous (for example, all real numbers between two real numbers), respectively. Such a working definition usually suffices for the needs of mathematics. See ANALYTIC GEOMETRY; NUMBER THEORY; PARAMETRIC EQUATION. [L.M.BL.]

Variable star

A star that has a detectable change in its intensity, often accompanied by other physical changes. The changes in brightness of a variable star may be a few thousandths of a magnitude up to 20 magnitudes or even more. The cause of the variation may be either an eclipsing effect caused by the motions within a system o[illegible] r more bodies

or an actual physical change within the star itself or in its surrounding atmosphere The physical or intrinsic variables will be considered here. For a discussion of the extrinsic or eclipsing variables, see BINARY STARS.

Nomenclature. The nomenclature of variable stars is based on the Argelander system. In this system, the first variable star in a constellation is assigned the letter R, the next S, and so on until Z; then the letters are repeated RR, RS, RZ

V335, V336, . These letters are combined with the genitive of the Latin constellation names and often used with the convenient six figure designation, the first four numbers of which give the hours and minutes of right ascension, and the last two the degrees of declination, for the epoch 1900 For example, 233815 R Aquarii (*south declination*), 213843 SS Cygni, 061115 CZ Orionis, and 184300 V603 Aquilae

A committee appointed by the International Astronomical Union decides when a variable should be accepted as such and assigns its name The second edition of the *General Catalogue of Variable Stars* lists 14,708 variable stars in our galaxy Undoubtedly there are many thousands of undiscovered variable stars in our galaxy, bright enough to be observed and as many thousands more too distant and too faint to be seen with present equipment

Importance. Studies of the variations of stars add to the general knowledge of the structure of the universe They are the most useful means available for the determination of distances and dimensions of our galaxy and of other galaxies Most of the variables show peculiarities in their variations when they are observed over long periods of time This necessitates long range programs for continuing observations over many years

Programs for the systematic visual observation of variables with relatively large ranges are carried out by groups of observers, such as the American Association of Variable Star Observers (AAVSO) These groups are composed mostly of amateur astronomers with their own equipment who derive much satisfaction from the knowledge that their conscientious work is of great value to scientific research Members of the AAVSO have contributed more than 1 500 000 observations in the nearly 50 years of its existance

Professional astronomers at the great observatories usually concentrate on observations of specific variables. Their observations are made visually and photographically, or, in the case of variables with small or rapid changes, with photoelectric photometers (see PHOTOMETRY) An example of a large photographic program is the Harvard Observatory photographic patrol which covered the sky for more than 50 years This collection of nearly 500 000 plates contains material for the study of variables brighter than twelfth magnitude.

Classification. Most intrinsic variables are either pulsating or eruptive (see accompanying table). They may be separated into three main divisions— those which vary with periodic regularity, those with cyclic or semiregular changes, and those which are completely erratic, or irregular. A few variables are unique and cannot be placed in any of these classes, and others seem to be combinations of several classes

Representative types of intrinsic variables

Class	Range of period or cycle	Color and luminosity	Amplitude (magnitude)	Number*
Pulsating				
RR Lyrae type	0 05–1 2 days	White	<2	2126
Classical cepheid	1–70 days	High luminosity yellow	0 1–2	610
Long-period	80–1000 days	Red giants	2.5–6+	3637
Semiregular	30–1000 days	Red giants	<1 5–2	1675
Others				1487
				9855
Eruptive				
U Gem and Z Cam	10–600 days	Subdwarfs	2–6	127
Flare stars	?	Red dwarfs	1–6	13
Recurrent novae	20 years–?	Dwarfs?	7–9	6
Novae	centuries?	Dwarfs?	8–16	140
Supernovae	?	?	20+	7
Nebular variables	?	Yellow dwarfs		590
Others				74
				959

* From B V Kukarkin and P P Parenago *General Catalogue of Variable Stars* 2d ed 1958 In addition there are listed 2760 eclipsing stars, 982 unstudied variables and 152 others unique or constant

The light curve of a variable shows how the brightness changes with time Observed magnitudes are plotted against a time scale, from which the times of maxima and minima may be determined An inspection of the light curve may also indicate if the variable is periodic or cyclic and gives a first approximation of the period Furthermore, the shape of the curve, that is, the steepness of the rise and fall, together with the time between maxima, may give a clue to the type of variability (Fig 1)

The spectra of variable stars cover the whole sequence of spectral classes, from the hot, highly luminous blue stars at one end, to the cool, red stars at the other (see STAR) Stars of some types of variability may be recognized as variables by means of their spectra Specifically, a star with bands of metallic oxides or carbon, and with emission lines of hydrogen, is almost certainly a long-period variable or possibly a semiregular variable Many fainter novae are discovered from their spectra, which at one stage of their development show many emission lines, including the so-called "forbidden lines" of oxygen, neon, iron and other elements (see ASTRONOMICAL SPECTROSCOPY)

Periodic variables are divided into two main classes: cepheids and long-period, or Mira-type variables There are many intermediate examples, so that no definite dividing lines may be drawn

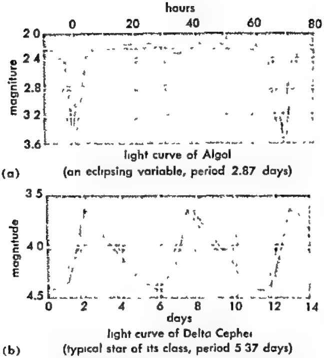


Fig. 1. Comparison of (a) eclipsing variable stars and (b) intrinsic variable stars

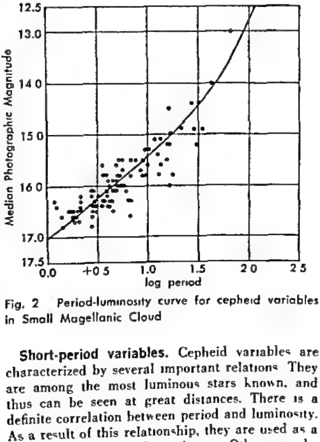


Fig. 2. Period-luminosity curve for cepheid variables in Small Magellanic Cloud

Short-period variables. Cepheid variables are characterized by several important relations. They are among the most luminous stars known, and thus can be seen at great distances. There is a definite correlation between period and luminosity. As a result of this relationship, they are used as a measuring tape for the universe. Other correlations exist between period and spectrum (temperature), period and radial velocity, and period and form of light curve.

Both spectrum and radial velocity of a cepheid change as the brightness changes. The spectrum at maximum is about a whole class earlier than at minimum. The greatest velocity of approach (negative velocity) occurs at maximum light and greatest recession velocity (positive) at minimum light.

Classical cepheids are population I stars and are situated within about 1 kiloparsec of the galactic plane.

The period-luminosity relation was first noticed in 1910 by Henrietta Leavitt at Harvard in the course of her work on the cepheids in the Magellanic Clouds and was further developed by Harlow Shapley (Fig. 2).

If the zero point of absolute magnitude could be established for cepheids of any given period, the distances of any stellar system in the universe, in which cepheids are found, could be determined from

$$M = m + 5 - 5 \log D$$

where M is absolute magnitude, m is apparent magnitude, and D is distance in parsecs. In 1950 the zero point was finally established by Walter Baade at Mount Wilson and Palomar Observatories, from his work on the cepheids in the Andromeda nebula.

Intrinsic variables with the shortest known periods are the cluster-type cepheids or RR Lyrae stars. RR Lyrae itself is one of the brightest and best studied of the group, with a period of 13.5 hours and a range of brightness from magnitude 6.9 to 8.0. The shortest period known, 89 minutes, is that of CY Aquarii.

The RR Lyrae stars are fainter than the classical cepheids. They have an absolute photographic magnitude of 0.0 and show no progression of luminosity with period. Therefore, if an RR Lyrae star can be identified in a cluster or galaxy, and its apparent magnitude determined, the distance of that system may be obtained from the relationship stated above.

RR Lyrae stars do not form an extension of the period-luminosity curve of the classical cepheids but lie in a branch about one and one-half magnitudes fainter. They are population II stars and form a spheroid in our galaxy, concentrated toward the galactic nucleus.

A group of peculiar cepheids, the W Virginis stars, are closely related to the RR Lyrae stars. They are also population II stars, have high velocities, and have periods greater than 10 days. A further continuation of this branch of the period-luminosity curve seems to be the RV Tauri stars, with periods between 40 and 150 days. Their spectra are more similar to RR Lyrae and W Virginis stars than to the classical cepheids.

The pulsation theory has been used to explain the variation of cepheids and related variables. The theory was first developed by Shapley (1914), by Arthur S. Eddington (1919), and later by many others. According to this theory, the star goes through regular cycles of expansion and contraction. As it expands, the pressure in its interior decreases, the temperature goes down, and it becomes redder and fainter. When the star reaches its maximum diameter, the process is reversed; it begins to contract, pressure increases, temperature rises, and it becomes bluer and brighter.

A simple process of expansion and contraction of the star as a whole fails to explain some of the observed peculiarities It is assumed that when the interior of the star pulsates, the compression waves pass through the lower levels of the star's atmosphere and reach the higher levels at some later time In this way, the observed lag of greatest velocity of approach relative to maximum brightness may be explained.

Long-period variables. Long-period or Mira-type variables are by far the most numerous of the known variables in our galactic system. They are red giants with spectra of either class M or one of the carbon types, and usually have hydrogen emission lines The spectra are of later type (redder and cooler) at minimum than at maximum

A study of the spectra of long period variables provides information about the layering of the atmosphere of a star The bright lines of hydrogen appear to be produced near the surface of the star, and the metallic vapors form a great cloud floating above the surface. The bright lines found in most stellar spectra originate in the distended atmospheres of high-temperature stars See STELLAR EVOLUTION

Several hundred long-period variables are brighter than tenth magnitude at maximum, and their large variations make them spectacular stars to observe with the unaided eye and small telescope Many of them have been observed for more than 100 years. Mean light curves and mean periods have been determined for most of the brighter ones, and many of them show decided variations in the shape of light curve and length of period from epoch to epoch The greatest irregularities occur in the heights and shapes of maxima For a discussion of the most famous long-period variable, see MIRA.

Many long-period variables are high-velocity stars and belong to population II, but some of the longer-period ones are more concentrated toward the galactic plane and have population I characteristics Others seem to form a link between the two populations

Long period variables may be caused by the same action as the cepheid variables Their great size and lower density could account for their irregularities

Irregular variables. Nonperiodic variables may be separated to some extent by their spectra Semiregular red giants are the most numerous and are closely related to the long-period variables Amplitudes are small ranging from less than 2 magnitudes to infinitesimal amounts Possibly all red stars with molecular bands in their spectra are variable

Most red variables with periods less than 150 days are subject to great irregularities, and some lose all semblance of periodicity at times, others are completely irregular

The most violent explosions or eruptions occur in the novae and supernovae Increases of 10–12 magnitudes have been observed within a few hours Previous to the explosion, novae are probably dense blue dwarfs (see NOVA; SUPERNOVA).

Other stars which have rapid increases of brightness are the U Geminorum and Z Camelopardalis stars (eruptive stars in the table). These also are blue stars with dwarf characteristics and are often referred to as dwarf novae. Their eruptions appear at semiperiodic intervals. A relationship exists between the amplitude of the variation and the duration between eruptions; variables with greater ranges have longer periods.

A small group of novae, called recurrent novae, have been ob[illegible]

Flare stars are red dwarfs, with emission line spectra, which more than double their brightness in a few minutes. Little is known about the frequency of the eruptions, and most of them have been found by accident Variable-star observers are making special studies of a group of red dwarfs to accumulate frequency statistics of flares.

The best known of the flare stars is UV Ceti, the typical star of its class It is normally of the thirteenth magnitude and has been observed to increase more than 6 magnitudes in less than 1 minute The decrease is almost as rapid. Many lesser flares have been observed visually, spectroscopically, photographically, and photoelectrically.

Nebular variables, also known as T Tauri, RW Aurigae, or Orion variables, are subject to rapid changes at irregular and unpredictable intervals. Most of them are dwarf stars with absorption spectra of classes G to M; in addition, many of them have peculiar bright lines, especially the H and K lines of calcium A few of them, such as R Monocerotis and R Coronae Australis are definitely associated with funnel-shaped nebulae that also vary. [M W.M.]

Bibliography: L. Campbell and L. Jacchia, *Story of Variable Stars*, reprint, 1945; B. V. Kukarkin and P. P. Parenago, *General Catalogue of Variable Stars*, 2d ed., 1958; C. Payne-Gaposchkin, *The Galactic Novae*, 1957

Variometer

A geomagnetic device for detecting and indicating changes in one of the components of the magnetic field vector—usually magnetic declination, the horizontal intensity component, or the vertical intensity component. When used with a suitable recording unit, a set of three variometers form[illegible]

[illegible] permanent bar magnet, ordinarily about 1 cm long, suspended with a plane mirror from a fine quartz fiber 5–15 cm in length A lens fixed in front of the mirror serves to focus to a point a beam of light reflected from the mirror to recording paper mounted on a rotating drum With no torsion in the fiber, the magnetic axis of the magnet will be

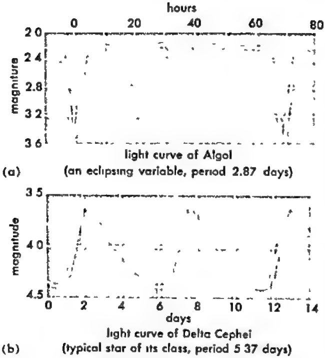


Fig. 1 Comparison of (a) eclipsing variable stars and (b) intrinsic variable stars

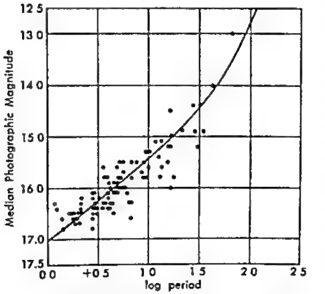


Fig. 2 Period-luminosity curve for cepheid variables in Small Magellanic Cloud

Short-period variables. Cepheid variables are characterized by several important relations. They are among the most luminous stars known, and thus can be seen at great distances. There is a definite correlation between period and luminosity. As a result of this relationship, they are used as a measuring tape for the universe. Other correlations exist between period and spectrum (temperature), period and radial velocity, and period and form of light curve.

Both spectrum and radial velocity of a cepheid change as the brightness changes. The spectrum at maximum is about a whole class earlier than at minimum. The greatest velocity of approach (negative velocity) occurs at maximum light and greatest recession velocity (positive) at minimum light.

Classical cepheids are population I stars and are situated within about 1 kiloparsec of the galactic plane.

The period-luminosity relation was first noticed in 1910 by Henrietta Leavitt at Harvard in the course of her work on the cepheids in the Magellanic Clouds and was further developed by Harlow Shapley (Fig. 2).

If the zero point of absolute magnitude could be established for cepheids of any given period, the distances of any stellar system in the universe, in which cepheids are found, could be determined from

$$M = m + 5 - 5 \log D$$

where M is absolute magnitude, m is apparent magnitude, and D is distance in parsecs. In 1950, the zero point was finally established by Walter Baade at Mount Wilson and Palomar Observatories, from his work on the cepheids in the Andromeda nebula.

Intrinsic variables with the shortest known periods are the cluster-type cepheids or RR Lyrae stars. RR Lyrae itself is one of the brightest and best studied of the group, with a period of 13.5 hours and a range of brightness from magnitude 6.9 to 8.0. The shortest period known, 89 minutes, is that of CY Aquarii.

The RR Lyrae stars are fainter than the classical cepheids. They have an absolute photographic magnitude of 0.0 and show no progression of luminosity with period. Therefore, if an RR Lyrae star can be identified in a cluster or galaxy, and its apparent magnitude determined, the distance of that system may be obtained from the relationship stated above.

RR Lyrae stars do not form an extension of the period-luminosity curve of the classical cepheids but lie in a branch about one and one-half magnitudes fainter. They are population II stars and form a spheroid in our galaxy, concentrated toward the galactic nucleus.

A group of peculiar cepheids, the W Virginis stars, are closely related to the RR Lyrae stars. They are also population II stars, have high velocities, and have periods greater than 10 days. A further continuation of this branch of the period-luminosity curve seems to be the RV Tauri stars, with periods between 40 and 150 days. Their spectra are more similar to RR Lyrae and W Virginis stars than to the classical cepheids.

The pulsation theory has been used to explain the variation of cepheids and related variables. The theory was first developed by Shapley (1914), by Arthur S. Eddington (1919), and later by many others. According to this theory, the star goes through regular cycles of expansion and contraction. As it expands, the pressure in its interior decreases, the temperature goes down, and it becomes redder and fainter. When the star reaches its maximum diameter, the process is reversed, it begins to contract, pressure increases, temperature rises, and it becomes bluer and brighter.

erated with an optical distance of about 175 cm between variometer lens and recording drum. The standard recording distances for H and Z variometers are 125 and 225 cm, respectively, although this can be varied to suit the design of the magnetograph. Sensitivity of the H variometer is a function of fiber stiffness (torsion constant) and the strength of the recording magnet, it varies from about 2 γ/mm ($1\ \gamma = 10^{-5}$ oersted) of spot displacement for low-latitude normal installations to about 40 γ/mm for high-latitude storm magnetographs, which are capable of recording the full range of severe magnetic storms. Similarly, the sensitivity of the Z variometer is a function of the strength of the magnet and the location of its center of mass relative to the line of knife edges, the range of sensitivities is from about 3 γ/mm to perhaps 40 γ/mm

When the three variometers are installed to record on one strip of paper, their proximity to one another results in some interaction among the various fixed and suspended magnets, resulting usually in slight misorientation of the latter. This is easily corrected in the adjustments at installation and when this is properly done, the effect on the records is negligible

The components to which variometers may be designed to respond are by no means limited to D, H, and Z, for they may be arranged to record ... the total ... ther than the permanent-magnet type are widely used; for some examples, see MAGNETOMETER [J H NE]

Bibliography. See MAGNETOMETER.

Varistor

Any variable resistor whose resistance depends upon voltage current, or polarity. All semiconducting diodes are varistors. Varistors can have either a symmetrical or a nonsymmetrical current voltage relation.

The most common symmetrical varistor material is silicon carbide, sold by certain companies under the trade name Thyrite. In varistor production silicon carbide is crushed, mixed with a small amount of graphite and a ceramic binder, and then fired in a furnace to form a ceramic matrix. Electrodes are applied by spraying or by chemical reduction of painted-on layers of metal organic paints

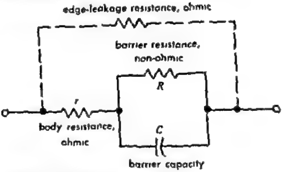
Other symmetrical varistors can be made of semiconducting material in which alternate layers of p and n type material are formed in such a way that there are an even number of p-n junctions in the material (see JUNCTION DIODE). These devices are commonly called stacked rectifiers

Nonsymmetrical varistors can be formed by a single p-n junction existing in a semiconducting material, by point-contact rectifiers or by special treatment of copper or selenium to form a copper-oxide or selenium rectifier

Copper-oxide rectifiers are formed by oxidizing copper disks or washers at 1000°C to form cuprous ... and the con- ... base is ... nent an ... silver is electrode usually of ... placed on top of the copper oxide to form a completed varistor.

In forming the selenium rectifier, a thin layer of elemental selenium is applied to one side of a base plate usually made of nickel, nickel-plated iron, or aluminum. The unit is then heat treated in order to change the selenium from the amorphous form to a crystalline form that is semiconducting. Then an electrode of cadmium or cadmium bearing alloy is sprayed on the surface of the selenium. The contact between the selenium and cadmium is rectifying, while the contact between the base plate and selenium is ohmic

The equivalent circuit of a varistor is shown in the illustration



Equivalent circuit of a varistor

It appears that, as the techniques are improved for producing semiconducting rectifiers composed of p-n junctions, these devices will take over most of the applications for varistors. See SEMICONDUCTOR RECTIFIER [C.L H]

Varnish

A transparent surface coating which is applied as a liquid and then changes to a hard solid. Varnishes are solutions of resinous materials in a solvent, which may dry by the evaporation of the solvent or by a chemical reaction, either with oxygen from the air or by some other means.

Spirit varnishes are those in which the evaporation of solvent is the only drying process. In this case the solvent is usually alcohol, although the term is used for similar coatings made with other solvents. Shellac varnish, made by dissolving shellac in alcohol, is the most common of this type. Oleoresinous varnishes are made by treating a drying oil with a resin usually with heat, and dissolving the reaction product in a solvent, usually a petroleum fraction. Drying of these varnishes results from the evaporation of the solvent, followed by polymerization of the drying oil portion, a reaction which is accelerated by metallic driers added to the varnish. For a discussion of the mechanism of ...

aligned in the magnetic meridian. As the meridian changes its direction, because of normal daily variations or magnetic storm disturbances, the magnet follows.

Horizontal intensity variometer. Although a larger, stiffer fiber is used, the H variometer is essentially the same as the D instrument. Enough torsion in the fiber is introduced at the time of installation, by twisting its upper end, to cause the magnet to turn 90° out of the magnetic meridian. At the sensitivity normally used, this adjustment results in a stable condition. The magnet responds to an increase of H by turning about the suspension axis. In order that the suspended magnet may not respond appreciably to changes in declination, it is aligned with the magnetic prime vertical to within about 0.5° for the standard magnetic observatory.

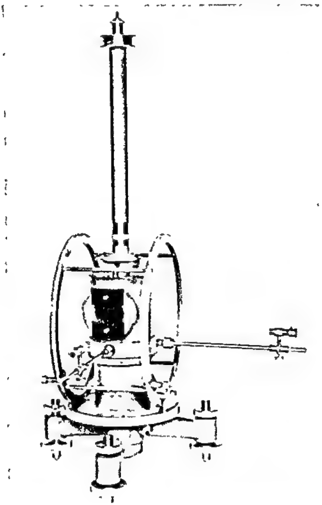


Fig. 1. Horizontal intensity (H) or declination (D) variometer equipped with Helmholtz coil for calibration. (U.S. Coast and Geodetic Survey)

Vertical intensity variometer. Employing a larger permanent magnet because of mechanical construction requirements, the Z variometer is equipped with very fine steel knife edges or pivots resting on agate planes or saddles and balanced so that its magnetic axis is horizontal. In this case, the torque due to the reaction of the magnetic moment with the Z component of the field is balanced against a torque due to the action of gravity on the laterally displaced center of mass of the magnet, and the magnet tilts one way or the other when the vertical field changes. A mirror mounted on the magnet indicates its movement. Suspension principles using other than the knife edges, for example, two horizontally stretched quartz fibers, have been introduced by some makers.

Both the H and Z variometers balance the action of a magnetic moment against a mechanical couple. The change in strength of the magnet with temperature may be compensated in several ways: magnetically, by suitably placed permanent magnets mounted on the frame of the variometer, as most commonly used in the United States, or by using a compound magnet assembly that has a zero over-all temperature coefficient; optically, by leading the path of the light beam through a prism or mirror supported on a bimetallic strip; or mechanically (for the Z variometer), by an adjustable counterpoise on a threaded shaft of relatively high temperature coefficient of expansion. However, all magnetic observatories provide a large amount of thermal insulation in the walls of the variation room so that the temperature changes are kept to a very low rate, and most observatories in regions of cold climate are equipped with thermostatically controlled alternating-current electric heaters.

Sensitivity characteristics. D variometers of normal sensitivity (usually 1.0′ or 0.5′ of arc/mm of displacement of the recording light spot) are op-

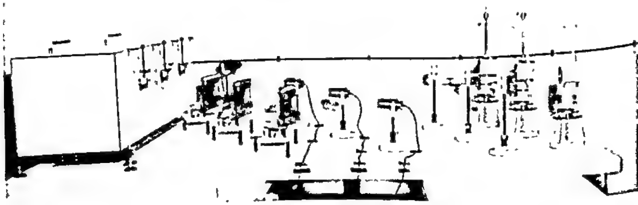


Fig. 2 Coast and Geodetic Survey magnetograph. (U.S. Coast and Geodetic Survey)

strands. In exceptional types of secondary growth, as in some monocotyledons, xylem and phloem may be combined into vascular bundles, which are imbedded in parenchyma or sclerenchyma; and in some dicotyledons wide rays are formed by the interfascicular cambium so that the secondary vascular tissues appear, like the primary, as a system of strands (see DICOTYLEDONEAE; MONOCOTYLEDONEAE; PARENCHYMA; SCLERENCHYMA).

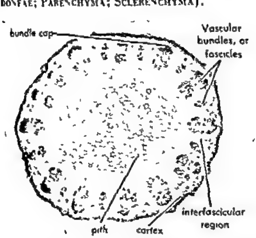


Fig 1. Cross section of *Helianthus* (sunflower) stem

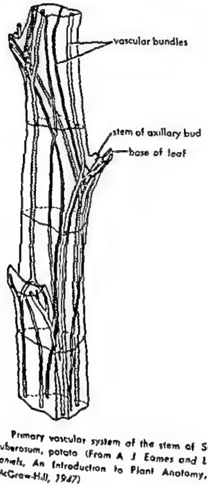


Fig 2 Primary vascular system of the stem of *Solanum tuberosum*, potato (From A J Eames and L H MacDaniels, An Introduction to Plant Anatomy, 2d ed, McGraw-Hill, 1947)

Types. The vascular bundles vary in structure, chiefly with reference to the arrangement of xylem and phloem. The most common type of bundle in seed plants is the collateral in which the phloem appears on one side of the xylem (Fig 3). The phloem may be rather deeply imbedded in the xylem in a collateral bundle; then the xylem assumes the shape of a V in cross sections (Fig 4). In a bicollateral vascular bundle the phloem occurs on both sides of the xylem (Fig 5).

The third main type of vascular bundle is the concentric, which occurs in two forms. In the amphivasal bundle (Fig. 6), the xylem surrounds the phloem, whereas in the amphicribral vascular bundle (Fig. 7), the phloem surrounds the xylem

The vascular bundles contain various amounts of supporting tissue, usually fibers. These may form a complete sheath enclosing the vascular tissues, or they may appear as strands outside the phloem or the xylem or on both sides As seen in cross sections, such strands are often called bundle caps

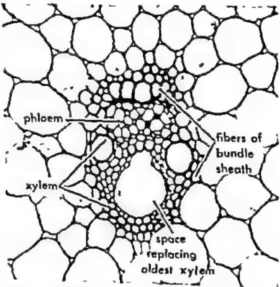


Fig 3 Cross section of a closed collateral vascular bundle from stem of *Zea mays*, maize (From K Esau, Plant Anatomy, Wiley, 1953)

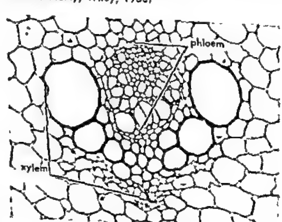


Fig 4 Cross section of a closed collateral vascular bundle from stem of *Asparagus*

in the appearance or the bulk properties Varnish coatings on wood are used to protect against abrasion, staining, and weather, and to reduce the penetration of water and other materials, without obscuring the grain or changing the color materially. Varnishes on metal, often called lacquers, are used to reduce corrosion without changing the metallic appearance. Varnishes are used on masonry to reduce the penetration of moisture and the damage from freezing. Paper is coated with varnish for moisture resistance, and to keep printing from being damaged.

Shellac varnishes are used where fast-drying, slightly colored coatings of low cost are needed, and where resistance to water, solvents, and weather is not required. Oleoresinous varnishes have greater resistance, but their color is usually darker. Spar varnishes are oleoresinous varnishes, often made from tung oil and a phenolic resin, which have good weather-resistance. They were originally used for marine finishes.

Varnishes are also used extensively as insulating coatings for wires and as vehicles for paints *See* INSULATION, ELECTRIC; PAINT.

Asphalt varnishes are made by treating a bituminous material, such as gilsonite, with a drying oil and dissolving the reaction product in a solvent These varnishes are black and opaque, unlike other varnishes They are used for insulation, and for metal coatings for heat- or corrosion-resistance.

Varnish stains contain a dye dissolved in the solvent; they are used to change the color of a wood without obscuring the grain. *See* SURFACE COATING. [F SD]

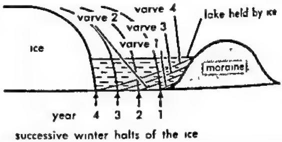
Varnish tree

This plant, *Rhus vernicifera*, also called lacquer tree, is a member of the sumac family (Anacardiaceae) and a native of China, but it has long been cultivated in Japan. When the bark is cut, it exudes a milky juice which darkens and thickens on exposure. This is the lacquer long used in China and Japan. When properly applied, the thin transparent film becomes a varnish of extreme hardness. Lacquer is a remarkably protective coating as it is not altered by acids, alkalies, alcohol, or heat up to 160°F. Nut galls, iron in solution, and gold or other metals are mixed with the lacquer before drying to make the various kinds of lacquers. The process of lacquering is technical and tedious, sometimes requiring 300–400 coats and several years to complete the finish of one item. *See* LACQUER, SAPINDALES [P D S.]

Varve

A distinctive, thin annual sedimentary layer, the lower part consisting of coarser, lighter-colored clay and silt that was deposited in summer, and the upper of a finer-grained, darker clay deposited in winter. Numerous successive varves, generally less than 1 in. thick, accumulated in temporary lakes near melting glaciers. Thicker and thinner varves at different places can be matched like tree rings.

Varved clay of glacial lake. Dark layers deposited in winter. *(Photograph by F. T. Thwaites)*



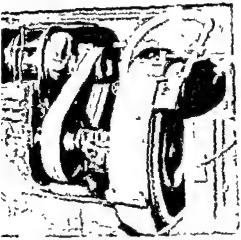
Formation of annual varves during ice-retreat from a moraine Each varve terminates at the line to which the ice receded each year *(From R N C. Bowen, The Exploration of Time, Philosophical Library, 1958)*

In Sweden, geologic history has been traced back about 18,000 years in this way. Similar layers occur in ancient rocks but it is difficult to determine whether the layers are annual; many of them represent longer or less regular sedimentary cycles. *See* SEDIMENTATION (GEOLOGY). [J.M W.]

Vascular bundles

Strandlike parts of the vascular system containing xylem and phloem (Fig. 1) The division of the vascular system into bundles is characteristic of stems, leaves, and flowers In these plant parts the bundles are interconnected into complex systems (Fig 2) In the root, however, the xylem and phloem are not combined into bundles but usually alternate with each other along different radii *See* FLOWER (BOTANY); LEAF (BOTANY), ROOT (BOTANY), STEM (BOTANY)

Occurrence in stems. Vascular bundles are *characteristic of stems in primary state of growth* During secondary growth xylem and phloem may also be formed between the bundles, that is, in the interfascicular regions (*see* PHLOEM, XYLEM) As a result of such growth the secondary vascular tissues form a continuous cylin[illegible] of a system of



2 Multiple V-belt drive (Goodyear Tire and Rub- Co)

wer to be transmitted exceeds the capacity of a igle belt (Fig 2) The driving and driven pulleys multiple drives carry separate grooves for each lt, the grooves being spaced to maintain a small tarance between the parallel belts To distribute e load evenly in a multiple drive the belts should matched, old and new belts should not be mixed e BELT DRIVE, PULLEY [RCF]

ectograph

A picture or drawing having self-contained light polarization At each point of such a picture, one can control the *direction* and magnitude of the polarization and the image can be expressed as a nonuniform vector field, hence the name vectograph Deep shadows have strongest polarization, while the highlights have none. A pair of stereoscopic vectograph pictures with axes of polarization at 90° to each other can be overlaid physically and yet can be separated optically and rendered selectively visible to each eye by analyzers. Spectacles made of two analyzers with axes at 90° to each other permit each eye to see the correct stereo picture, producing three-dimensional viewing No other equipment is necessary. Vectographs can be made in black and white, or color, and can be projected with an ordinary single lens projector. For direct viewing a reflection print is used Unlike an ordinary photograph, which is isotropic and has opaque silver or dye deposits in a gelatin film, a vectograph must have a dichroic dye deposited on a birefringent, highly oriented (stretched) film, for example, one made from polyvinyl alcohol A dichroic stain such as iodine can be employed for black and white prints *See* POLARIZED LIGHT; STEREOSCOPY. [J.MH.]

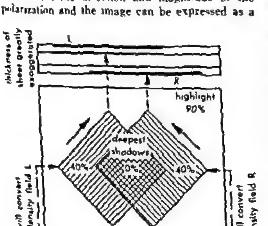


Diagram of superposed stereoscopic vectographs The solid arrows indicate axes of polarization, L, left eye image; R, right eye image, percentages refer to percentage of light transmission

Bibliography: E H. Land, *Opt. Soc. Am.*, *30*(6):230–238, *1940*, J Mahler, Photographic science and technique. *Phot Soc Am. ser* 2, 1(8): 84–87. 1954.

Vector (mathematics)

A directed line segment. As such, vectors have magnitude and direction Many physical quantities, for example, velocity, acceleration, and force, *are* vectors Vectors are widely used in *mathematical* physics *See* CALCULUS OF VECTORS.

Vectorcardiography

An electro... the re... *heartl*... as a ... by a negative point sink (*see* DIPOLE) At any instant of time during cardiac activity this hypothetical entity has magnitude and direction and can *therefore be* treated as a vector quantity (*see* CALCULUS OF VECTORS) By considering the surface electrocardiographic leads as vector components, size and direction of the cardiac vector forces can be obtained either by calculations (vector-electrocardiography) or by recording the resultant of *two vector* components on an oscilloscopic screen Combining the *image* of lead components obtained *from* three planes of the body, a three-dimensional *reconstruction* of the heart *vector* time course can be traced (spatial vectorcardiography)

To obtain relatively distortion-free scale... for ...

... represent most ... the *x, y, and z* planes of the body ... eral ...

... *x, y*, and *z* represent the orthogonal components of the central heart vector ($\overrightarrow{E}$), and

Fig 5. Cross section of an open bicollateral vascular bundle from stem of *Cucurbita*, squash. (From K. Esau, *Plant Anatomy*, Wiley, 1953)

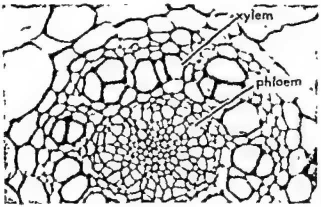


Fig. 6. Cross section of a concentric amphivasal vascular bundle from rhizome of *Acorus calamus*, sweet flag

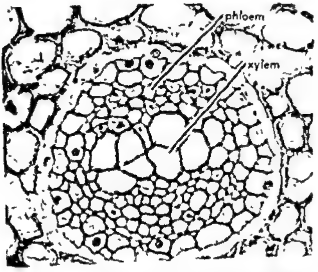


Fig. 7. Cross section of a concentric amphicribral vascular bundle from rhizome of *Polypodium*, a fern. (From K Esau, *Plant Anatomy*, Wiley, 1953)

Origin and development. The primary vascular bundles differentiate from a meristem called procambium or provascular tissue (*see* MERISTEM, APICAL). If all of this meristem becomes xylem and phloem, the resulting bundle is incapable of further growth and is commonly called closed vascular bundle. Such bundles are characteristic of lower vascular plants, monocotyledons, and those dicotyledons that have no secondary growth (*see* PLANT GROWTH). If secondary growth occurs, the last procambial cells become cells of the vascular cambium—fascicular cambium—which, together with the interfascicular cambium, forms the secondary xylem and phloem (*see* MERISTEM, LATERAL). Bundles capable of further growth are called open bundles. [K.E.]

Bibliography: *See* PLANT ANATOMY.

V-belt

A belt, usually endless, having a trapezoidal cross section and designed to run in a pulley or sheave with a V-shaped groove. V-belts are of rubber compounds, cotton cords and fabric (Fig. 1). They provide quiet, compact power transmission, absorb shocks, and operate with small initial tension and low bearing pressures. The nature of the wedging action of the belt in the pulley groove allows them to transmit considerably more power than a flat belt of the same width

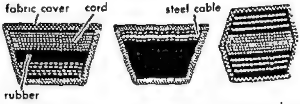


Fig 1 Cross sections of V-belts (Goodyear Tire and Rubber Co)

Only the angular sides of the V-belt should contact the sheave Because of this requirement, little adjustment is required to compensate for stretch and wear To ensure proper contact the belt should be of proper length to leave a minimum of 1/4-in clearance between the inner belt surface and bottom of the pulley (sheave) groove. The included angle of the belt is approximately 40° and the sides are slightly concave to allow for bulging under tension in the pulley groove The included angle of the pulley groove is less than that of the belt. Recommended center distances for V-belt drives are from not less than the diameter of the larger pulley to not more than the sum of the diameters of the two pulleys although larger and smaller distances may be used The velocity ratios are inversely proportional to the pitch diameters of the sheaves where the pitch diameter is approximately equal to the outside diameter of the sheave minus the belt thickness

Multiple V-belt drives, cons[illegible] several separate belts in parallel gr[illegible] when the

See separate articles on vegetables listed under heir common names [H J.C.]

Bibliography: See AGRICULTURAL SCIENCE (PLANT)

Vegetable ivory

The seed of the tagua palm, *Phytelephas macrocarpa*, of tropical America. Each drupelike fruit contains 6–9 bony seeds. The extremely hard endosperm of the seed is used as a substitute for ivory. Vegetable ivory can be carved and tooled to make buttons, chessmen, knobs, inlays, and various ornamental articles. The hard, white seeds (coquilla nuts) of the Brazilian palm, *Attalea funifera*, are often used as a substitute for vegetable ivory in making the same or similar articles. See PALMALES. [P D S.]

Vegetation zones

The more or less distinct horizontal belts of vegetation that occur, one above another, on high mountains. The altitudinal zones which are considered here are correlated with altitudinal gradients of climatic factors that, in turn, reflect altitudinal variations in the atmosphere. These graded factors are chiefly those of light, heat, precipitation, evaporation, and growing season. See ENVIRONMENT.

Light. With an increase in elevation above sea level, air becomes less dense and is not able to hold as much moisture per unit volume. The lessening of density results in a decrease of barometric pressure of about 1 in./1000 ft. With a greater dispersion of gas molecules and a lower concentration of water vapor, the heat absorbing power of the air is diminished. In addition, as altitude increases, the distance through the atmosphere that the sun's rays must travel to reach the earth diminishes. Therefore, more light and more heat energy reach a surface at a high elevation than reach a similar surface at a low elevation. It has been estimated that about 20% of solar radiation is reflected or absorbed by the atmosphere before it reaches sea level, whereas only 11% is lost before it reaches a 14,000-ft high peak. The difference in light intensity is not of sufficient importance to photosynthesis to be a critical ecological factor, but the heating of the soil may produce very high surface temperatures during the day. The heat is lost rapidly by reradiation at night, and because the air is not able to absorb appreciable quantities of heat, the nocturnal temperatures are relatively low. Measurements on Pike's Peak, Colorado, for example, occasionally indicate surface soil temperatures near 140°F at a time when the air temperature 5 ft above the soil is about 70°F. Nocturnal reradiation also prevents the heat from being conducted into the deeper soil layers. On Pike's Peak, the soil at a depth of 10 in. may be 85°F cooler than the surface soil.

Temperature. Air masses that encounter mountains are forced upward and over the barrier. The rising air expands and is cooled adiabatically at a rate that averages about 1°F decrease in temperature per 330 ft rise in altitude in summer, 1°F per 400 ft rise in winter. The thermic rate varies considerably because of moisture, wind, slope exposure, and other factors.

The reduction of temperature as altitude increases also brings about later frosts in the spring and earlier frosts in the autumn and thus shortens the frost-free or growing season. In Arizona, for example, a station at an elevation slightly above sea level has an average frost-free season of 322 days, a station at an elevation of 2660 ft has an average frost-free season of 246 days, and another station at an elevation of 9000 ft, Mount Lemmon, has a frost free season that averages only 122 days. However, there is a considerable variation in the length of the frost-free season from place to place in a given vegetation type. Also, temperature alone may not suffice to determine the growing season, because many low plants may be covered by snow for several days or even weeks after the air temperature becomes favorable. Thus, it appears that the absolute length of the season of favorable temperatures is not closely correlated with altitudinal changes in vegetation.

Precipitation and evaporation. A portion of the water vapor contained in rising air masses condenses as the air mass is cooled and may form clouds or fall as rain or snow. Thus, mountains are often referred to as islands of greater precipitation. In the mountains, the annual precipitation generally increases in amount with an increase in altitude, at least up to a certain elevation. Mountains subject to moist winds from a prevailing direction may have heavy precipitation on windward slopes and a rain shadow or dry belt on the lee side. If the mountains are high enough, there is generally a maximum of precipitation on the intermediate slopes and a gradual decrease in average precipitation from that point to the mountain crests. Furthermore, the lower temperatures at high altitudes result in a lessening of evaporation and transpiration, so that precipitation may be more effective at higher elevations. The rate of diminution of the evaporation rate caused by cooling is partly offset by the fact that the evaporating power of the air increases with altitude as atmospheric pressure decreases. Decreases of evaporation losses that result from [illegible]

e [illegible]

b [illegible]

t [illegible] increases in wind velocity.

The decreased atmospheric pressure at higher altitudes also is accompanied by reductions in the partial pressures of oxygen and carbon dioxide. Thus, the oxygen concentration in the soil atmosphere decreases with altitude and may become a limiting factor for some species. Likewise, the lower partial pressure of carbon dioxide may be limiting to photosynthesis in many species.

Environment and vegetation. From the preceding summary of altitudinal gradation of climatic phenomena, it can be seen that mountain environments are complex and highly variable. The condi-

a, b, and c are constant coefficients which define the deviation of any lead from orthogonality The constants in turn may be considered the orthogonal components of another hypothetical entity, the lead vector $(\vec{L})$. *See* ORTHOGONAL POLYNOMIALS.

Vectorcardiography has as yet contributed little to the diagnostic advance in electrocardiography, but has clarified the interrelationship of lead connections and has allowed a remarkably simplified approach to the understanding of the complicated electrical activity of the heart. *See* ELECTROCARDIOGRAPHY; ELECTROPHYSIOLOGY (HEART). [H.HE.]

Bibliography: G. E. Burch, J. A. Abildskov, and J. A. Cronvich, *Spatial Vectorcardiography*, 1953; H. C. Burger and J B van Milaan, *Brit Heart J.*, 8:157, 1946, 9:154, 1947; E Frank, *Circulation Research*, 2:258, 1954

Veering wind

A wind which changes direction in a clockwise sense; for example, a change from a southerly to a westerly direction The wind may veer gradually, over a period of hours, or abruptly, in a few minutes or seconds on passage of a wind-shift line The wind veers when a cyclone passes eastward on a path north of the observer Veering with height is usually found on the east side of a cyclone, or west of an anticyclone, with warm air to the south. In the southern hemisphere, the meaning in terms of cardinal directions is reversed *See* WIND-SHIFT LINE. [C W N]

Vega

One of the brightest stars in the sky, apparent magnitude 0.1. Vega, or α Lyrae is a normal main sequence star of spectral type A0, having an effective temperature near 10 000°K Its distance is 8 parsecs, its absolute magnitude is +0 5 (40 times as bright as the Sun), and its radius 4 times that of the Sun The spectrum shows mainly the strong Balmer lines of hydrogen *See* STAR

[J.L GR]

Vegetable growing

A branch of horticulture relating to the production of a number of herbaceous plants or plant parts commonly referred to as vegetables. No concise definition of the term is possible: for example, such so-called vegetables as tomatoes, corn, and beans are technically fruits. *See* FRUIT (BOTANY) The term olericulture, referring to vegetable production, is used occasionally.

Vegetables are excellent sources of minerals and vitamins (*see* MINERAL; VITAMIN). Wider recognition of this fact has resulted in greater per capita consumption of carrots, lettuce, sweet corn, and many other highly palatable crops. More than 40 vegetables are grown by commercial farmers and home gardeners for fresh consumption and for processing. (*See* FOOD ENGINEERING).

Home gardening. A large assortment of vegetables can be raised in all parts of the United States if adapted varieties are selected *and plantings are* made to coincide with favorable climatic periods. A home garden can reduce family food costs, provide vegetables of higher quality than is available in many stores, and serve as recreation

Commercial vegetable growing. The vegetable industry has undergone dynamic changes since World War II. As a result of vastly improved production, handling, and transportation techniques, a wide variety of fresh and processed vegetables from distant areas is available in all parts of the United States throughout the year.

In the past, the term market gardening referred to the production of vegetables for local markets, truck farming described the production of crops in large volume for shipment to distant markets Small acreages, low transportation costs, and a diversity of crops characterized the market garden operation, whereas low labor and land costs were typical of the distant shipping regions These distinctions are no longer valid.

In recent years rapid growth of chain stores and changes in consumer buying habits have resulted in a concentration of purchasing power in the hands of comparatively few produce buyers. Emphasis on a steady supply and large volume of uniformly packaged vegetables has favored the long-season areas having favorable climate and large acreages that usually are great distances from the centers of dense population Individual vegetable farms are generally large and heavily mechanized (*see* AGRICULTURAL MACHINERY). California, Florida, Texas, and Arizona accounted for 60% of the total fresh vegetable production in the United States in 1957 despite the higher transportation costs involved Smaller vegetable farms near cities are declining in importance but still compete by selling at roadside stands, on farmers' markets, or through group-marketing organizations

Consumption of canned and frozen vegetables is increasing The West Coast, Wisconsin, and Minnesota lead in production. In many areas crops for processing are grown by men who produce for the fresh market also, in other areas, processing vegetables are raised by grain or livestock farmers in addition to their other products. Dehydrated onions and potatoes are becoming more popular, although dehydration is currently not an important means of preserving vegetables.

Smaller acreages of a few crops are forced in special structures at times other than their normal season of growth, for example, tomatoes, radishes and cucumbers are produced in greenhouses *during* the winter in Ohio Michigan, and other northern states Washington and Michigan are centers for rhubarb forcing

Additional commercial vegetable growing activities include production of vegetable seeds and of young tomato, cabbage and similar plants for sale to other growers The plant-growing industry is well developed in Georgia and Texas, while seed production is confined mainly to Idaho, Washington, *Oregon and California*

See separate articles on vegetables listed under their common names. [H.J.C.]

Bibliography: See AGRICULTURAL SCIENCE (PLANT).

Vegetable ivory

The seed of the tagua palm, *Phytelephas macrocarpa*, of tropical America. Each drupelike fruit contains 6–9 bony seeds. The extremely hard endosperm of the seed is used as a substitute for ivory. Vegetable ivory can be carved and tooled to make buttons, chessmen, knobs, inlays, and various ornamental articles. The hard, white seeds (coquilla nuts) of the Brazilian palm, *Attalea funifera*, are often used as a substitute for vegetable ivory in making the same or similar articles. See PALMALES. [P.D.S.]

Vegetation zones

The more or less distinct horizontal belts of vegetation that occur, one above another, on high mountains. The altitudinal zones which are considered here are correlated with altitudinal gradients of climatic factors that, in turn, reflect altitudinal variations in the atmosphere. These graded factors are chiefly those of light, heat, precipitation, evaporation, and growing season. See ENVIRONMENT.

Light. With an increase in elevation above sea level, air becomes less dense and

as much

of dens

sure of

sion of

wa

dir

distance through the atmosphere that the sun's rays must travel to reach the earth diminishes. Therefore, more light and more heat energy reach a surface at a high elevation than reach a similar surface at a low elevation. It has been estimated that about 20% of solar radiation is reflected or absorbed by the atmosphere before it reaches sea level whereas only 11% is lost before it reaches a 11,000-ft high peak. The difference in light intensity is not of sufficient importance to photosynthesis to be a critical ecological factor, but the heating of the soil may produce very high surface temperatures during the day. The heat is lost rapidly by reradiation at night, and because the air is not able to absorb appreciable quantities of heat, the nocturnal temperatures are relatively low. Measurements on Pike's Peak, Colorado, for example, occasionally indicate surface soil temperatures near 140°F at a time when the air temperature 5 ft above the soil is about 70°F. Nocturnal reradiation also prevents the heat from being conducted into the deeper soil layers. On Pike's Peak the soil at a depth of 10 in. may be 85°F cooler than the surface soil.

Temperature. Air masses that encounter mountains are forced upward and over the barrier. The rising air expands and is cooled adiabatically at a rate that averages about 1°F decrease in temperature per 330 ft rise in altitude in summer, 1°F per 400 ft rise in winter. The thermic rate varies considerably because of moisture, wind, slope exposure, and other factors.

The reduction of temperature as altitude increases also brings about later frosts in the spring and earlier frosts in the autumn and thus shortens the frost-free or growing season. In Arizona, for example, a station at an elevation slightly above sea level has an average frost-free season of 322 days, a station at an elevation of 2660 ft has an average frost-free season of 246 days, and another station at an elevation of 9000 ft, Mount Lemmon, has a frost-free season that averages only 122 days. However, there is a considerable variation in the length of the frost-free season from place to place in a given vegetation type. Also, temperature alone may not suffice to determine the growing season, because many low plants may be covered by snow for several days or even weeks after the air temperature becomes favorable. Thus it appears that the absolute length of the season of favorable temperatures is not closely correlated with altitudinal changes in vegetation.

Precipitation and evaporation. A portion of the water vapor contained in rising air masses condenses as the air mass is cooled and may form clouds or fall as rain or snow. Thus, mountains are often referred to as islands of greater precipitation. In the mountains, the annual precipitation generally increases in amount with an increase in altitude, at least up to a certain elevation. Mountains subject to moist winds from a prevailing direction may have heavy precipitation on windward slopes and a rain shadow or dry belt on the lee side. If the mountains are high enough, there is generally a maximum of precipitation on the intermediate slopes and a gradual decrease in average precipitation from that point to the mountain crests. Furthermore, the lower temperatures at high altitudes result in a lessening of evaporation and transpiration, so that precipitation may be more effective at higher elevations. The rate of diminution of the evaporation rate caused by cooling is partly offset by the fact that the evaporating power of the air increases with altitude as atmospheric pressure decreases. Decreases of evaporation losses that result from cooling have been demonstrated by several workers. However, these losses are also offset by lower soil temperatures, which reduce water uptake, and by increases in wind velocity.

The decreased atmospheric pressure at higher altitudes also is accompanied by reductions in the partial pressures of oxygen and carbon dioxide. Thus, the oxygen concentration in the soil atmosphere decreases with altitude and may become a limiting factor for some species. Likewise, the lower partial pressure of carbon dioxide may be limiting to photosynthesis in many species.

Environment and vegetation. From the preceding summary of altitudinal gradation of climatic phenomena, it can be seen that mountain environments are complex and highly variable. The condi-

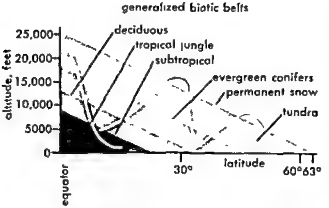


Fig. 1. Diagram to illustrate the variation of altitudinal position of vegetational zones with latitudinal change. *(From A. M Woodbury, Principles of General Ecology, McGraw-Hill, 1954)*

tions are never the same on two mountain masses, because many factors, such as slope gradient, exposure, direction of prevailing winds, nearness to oceans, geologic structure, and local topography, lead to individuality. In addition, plant life may respond to different limiting factors at various elevations. For example, studies in eastern Washington and northern Idaho have indicated that ecotones between vegetation types are related to critical deficiencies of heat at upper elevations, and to critical water shortages at lower elevations.

Generally, mountains in tropical, subtropical, and temperate areas are covered by forests, regardless of the nature of the surrounding vegetation The forests, often of several types, such as tropical broadleaf evergreen, temperate deciduous, and northern coniferous evergreen, are found at different elevations, and will extend upslope to a point at which they yield to a meadow growth composed of grasses, perennial herbs, and low shrubs. If the mountain is high enough, there may be an alpine tundra and an area of perennial snow without vegetation (Fig. 1). Even in widely separated areas with different floras, similar environmental conditions tend to support similar vegetation types. The variety of altitudinal zones thus increases with the height of the mountain, but it also increases with the nearness of the mountain to the Equator (Fig. 1). The latter statement is based on the fact that as one travels further from the Equator, he experiences climatic and vegetational changes parallel to those experienced when one climbs a mountain. Thus, fewer low elevation forms of vegetation are available in higher latitudes. Ultimately, in arctic regions, only tundra vegetation may occur on the mountains and in the antarctic there are many snow-covered ranges with no vegetation.

Altitudinal zonation. The altitudinal zonation of vegetation in the northern intermountain region of Idaho and eastern Washington can be taken as an example of the arrangement of zones in a temperate mountainous area (Fig. 2). In this area, the lower elevations are dominated by sagebrush, *Artemisia tridentata*. At higher elevations, several grassland types may occur and above these may be a poorly-developed woodland of juniper, *Juniperus scopulorum*, with or without limber pine, *Pinus flexilis* The ponderosa pine zone is the lowest of the well-formed conifer forests. It is often composed entirely of ponderosa pine, *P. ponderosa*, with an undergrowth of prairie grasses and forbs or shrubs. Douglas fir, *Pseudotsuga taxifolia*, occupies the zone above the ponderosa pine, but is often found in mixed stands with the pine or with species from the arbor vitae–hemlock zone that oc-

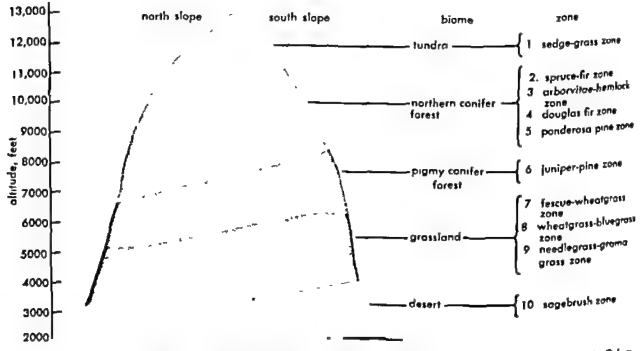


Fig 2. Zonation of vegetation in the northern intermountain region of *Idaho and eastern Washington.* *(Based on data by R. F. Daubenmire from E Odum, Fundamentals of Ecology, Saunders, 2d ed, 1959)*

curs above. The highest coniferous forest zone is the spruce-fir zone in which the Engelmann spruce, *Picea engelmanni*; alpine fir, *Abies lasiocarpa*; and mountain hemlock, *Tsuga heterophylla*, occur in various mixtures. The sedge-grass zone, or tundra, occupies all of the areas above the limits of forest growth.

Topography. Most mountainous areas are composed of a variety of geological formations, are interrupted by deep canyons, ravines, or passes, and have slopes with various exposures. In such areas, vegetation zones, therefore, generally are very irregular and may have projections or outliers that extend hundreds of feet in elevation above or below the main body. Usually, a given zone is higher on a warm, Equator-facing slope than on a cooler, pole-facing slope. Measurements in several areas have shown that temperatures on an Equator facing slope are usually a few degrees warmer than those on a pole-facing slope due to the distribution of solar energy on the slopes.

In narrow valleys, ravines, or canyons, inversions of temperature may occur frequently as masses of cold, more dense air flow down the valley walls and along the valley bottom. These cold air masses often accumulate to some depth in places where their paths are obstructed by constrictions in the valley or by other barriers. Thus, the minimum temperatures increase upslope from the valley bottom to a point at which the inversion ends, sometimes as high as 300 m above the valley floor, then they decrease in accord with the normal thermic gradient. As a response to the lower temperatures and greater moisture in such valleys, they are often occupied by pennantlike projections of vegetation zones that occur at much higher elevations on the mountain slopes. Thus, inversions of vegetation zones can often be seen in narrow valleys. In the northern intermountain region, for example, the bottom of a cool canyon may be occupied by a spruce-fir forest, the lower slope may be covered with [illegible] pine. Thus, the *ponderosa* pine zone is higher in altitude in such localities than is the spruce-fir zone. See CLIMAX COMMUNITY; PLANT COMMUNITY; TERRESTRIAL ECOSYSTEM. [J.S.M.]

Bibliography. R. F. Daubenmire, Vegetational zonation in the Rocky Mountains, *Botan. Rev.*, 9(6):325–393, 1943; R. F. Daubenmire, The life-zone problem in the *northern* intermountain region, *Northwest Sci.*, 20(2):28–38, 1946; F. Shreve, *The Vegetation of a Desert Mountain Range as Conditioned by Climatic Factors*, Carnegie Inst. Washington, Publ. 217, 1915.

Vegetation zones (world)

The areas of distribution on the continents of the four great plant formations: forests, grasslands, desert shrub and waste, and tundra. These are subdivided into more specific associations which have significance to human and animal life. For discussions of the compositions of the formations see DESERT; FOREST VEGETATION; PRAIRIE; STEPPE; TUNDRA. The relationship of vegetation to climate is so close that climatic classifications are commonly based on vegetation distribution, and the patterns of vegetation and climate when expressed on maps are remarkably similar. In general, humid lands are forested, arid lands bear desert shrub or are waste, and intermediate areas of semiaridity or marked seasonal drought are occupied by grasslands. The brief cool growing season of the Arctic Ocean coasts produces only tundra. There are numerous exceptions to these generalizations. Only in the tundra and forest lands near the North Pole is vegetation distributed in continuous bands or zones. Farther south, the vegetation belts are discontinuous, interrupted by rainfall and altitude differences.

Desert distribution. Xerophytic, or drought resisting, desert shrub or wasteland occurs on the continental west coasts between the latitudes of 20 and 30 in both hemispheres. From these situations it extends inland and toward the poles in discontinuous distributions. In North America, deserts occupy Lower California and much of northern Mexico and extend into the intermountain area of the United States. In the combined continent of Africa-Europe-Asia the desert belt, broken by mountains and highlands, reaches from the western Sahara to inner Asia. In the Southern Hemisphere the deserts are less extensive because of the tapering of the continents. The desert coastal strip of Chile and Peru is continued east of the Andes in Patagonia. The Kalahari and Namib deserts occupy parts of southwest Africa and Bechuanaland. Most of the central and western part of the Australian continent is desert.

Low-latitude vegetation distribution. The tropical forests and grasslands occupy a belt some twenty degrees wide on either side of the Equator. The rain forest occurs on the continuously humid lowlands of the Amazon basin, the Congo basin and adjoining coastal parts of Africa, eastern Madagascar, the East Indies including the Philippines and New Guinea, and coastal strips in India and southeast Asia. Wherever there is an appreciable drought in the climate the rain forest is supplanted by other types. In monsoonal and short drought situations this is the tropical semideciduous forest which occurs in eastern India, Indonesia, the east coasts of Brazil and south Africa, the northern coasts of Australia, and the west coasts of Central America. Where the drought is long or severe, grassy or thorny scrub forests or tropical grass-tree savannas occupy the land. These types are transitional between the rain forest and the tropical sides of the deserts. Their detailed distribution is complex and incompletely known. They occur in the Venezuelan highlands and the Brazilian plateau of South America. In Africa a great crescent of scrub forests and savannas surrounds the Congo basin extending from the Sahara margins in the north, including all of east Africa and south Africa.

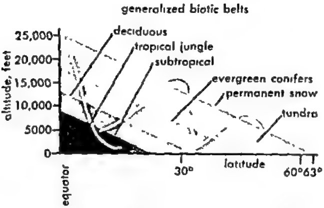


Fig. 1. *Diagram to illustrate the variation of altitudinal position of vegetational zones with latitudinal change. (From A. M Woodbury, Principles of General Ecology, McGraw-Hill, 1954)*

tions are never the same on two mountain masses, because many factors, such as slope gradient, exposure, direction of prevailing winds, nearness to oceans, geologic structure, and local topography, lead to individuality. In addition, plant life may respond to different limiting factors at various elevations For example, studies in eastern Washington and northern Idaho have indicated that ecotones between vegetation types are related to critical deficiencies of heat at upper elevations, and to critical water shortages at lower elevations.

Generally, mountains in tropical, subtropical, and temperate areas are covered by forests, regardless of the nature of the surrounding vegetation. The forests, often of several types, such as tropical broadleaf evergreen, temperate deciduous, and northern coniferous evergreen, are found at different elevations, and will extend upslope to a point at which they yield to a meadow growth composed of grasses, perennial herbs, and low shrubs. If the mountain is high enough, there may be an alpine tundra and an area of perennial snow without vegetation (Fig. 1). Even in widely separated areas with different floras, similar environmental conditions tend to support similar vegetation types. The variety of altitudinal zones thus increases with the height of the mountain, but it also increases with the nearness of the mountain to the Equator (Fig. 1). The latter statement is based on the fact that as one travels further from the Equator, he experiences climatic and vegetational changes parallel to those experienced when one climbs a mountain. Thus, fewer low elevation forms of vegetation are available in higher latitudes. Ultimately, in arctic regions, only tundra vegetation may occur on the mountains and in the antarctic there are many snow-covered ranges with no vegetation.

Altitudinal zonation. The altitudinal zonation of vegetation in the northern intermountain region of Idaho and eastern Washington can be taken as an example of the arrangement of zones in a temperate mountainous area (Fig. 2). In this area, the lower elevations are dominated by sagebrush, *Artemisia tridentata* At higher elevations, several grassland types may occur and above these may be a poorly-developed woodland of juniper, *Juniperus scopulorum*, with or without limber pine, *Pinus flexilis*. The ponderosa pine zone is the lowest of the well-formed conifer forests. It is often composed entirely of ponderosa pine, *P. ponderosa*, with an undergrowth of prairie grasses and forbs or shrubs. Douglas fir, *Pseudotsuga taxifolia*, occupies the zone above the ponderosa pine, but is often found in mixed stands with the pine or with species from the arbor vitae–hemlock zone that oc

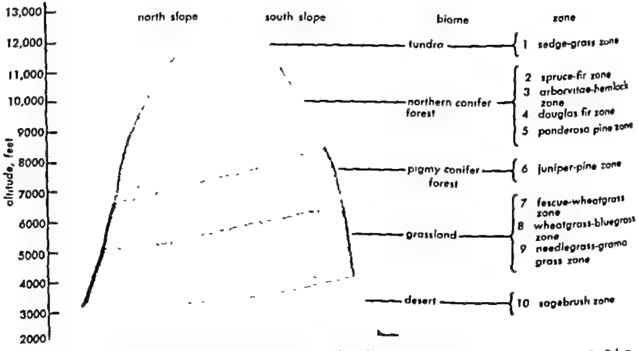


Fig. 2. *Zonation of vegetation in the northern Intermountain region of Idaho and eastern Washington. (Based on data by R. F Daube m E Odum, Fundamentals of Ecology, Sau 59)*

Rocky Mountains and the forest margins in central United States and southern Canada where their prairie parts have been almost completely eliminated by grain agriculture. In Africa-Europe-Asia the mid latitude grasslands begin at the Atlas mountains and extend to Manchuria. This band encloses the Black and Caspian Seas and the Euro-Asiatic deserts In the Southern Hemisphere the grasslands are smaller and discontinuous. The Pampas of Argentina and Uruguay is mostly prairie, as is the Veld of southeastern Africa In Australia the grasslands of the southeast bear also a scanty growth of small broadleaf evergreen trees.

Coniferous forest and tundra distributions. The northern coniferous, or boreal, forest extends in a wide circumpolar belt across the northern parts of Eurasia and North America between the latitudes of approximately 50 and 70°. Other coniferous forest types commonly of larger trees extend southward along the [illegible]

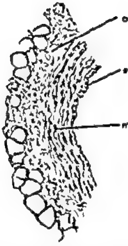
The vegetation pattern of the world is completed by the tundra association which occupies the narrow coastal belt between [illegible]

[illegible] [C.M.D.]

Bibliography V. C. Finch, G. T. Trewartha, A H Robinson, and E H Hammond, *Elements of Geography*, 4th ed., 1957, P E James, *A Geography of Man*, 1951

Vein

The relatively thin-walled blood vessels that [illegible] blood from [illegible] several cha[illegible] from their [illegible] the arteries They are much more variable in location and distribution, they are usually larger than the corresponding artery as a result of a slower rate of blood flow and other factors Veins have thinner less muscular walls and many veins contain valves which prevent backflow of blood particularly in areas where gravity exerts a large effect Anastomoses occur more frequently in the veins than they do in the arteries As a result, the venous system has a more netlike pattern See ARTERY.



Portion of cross section through common digital vein of a man. (From A. A. Maximow and W. Bloom, *A Textbook of Histology*, 6th ed., Saunders, 1953)

The walls of veins usually have three layers, although the smaller venules may appear much like the single walled capillaries. There is an outer connective tissue covering, the adventitia, a middle muscular coat which is less developed than that of arteries, and an inner endothelial layer continuous with the lining of the remainder of the circulatory system

The venules, which receive blood from the capillaries, gradually merge to form larger and larger veins. As is the case with the arterial system, two major circuits, the pulmonary and the systemic, are recognized The former carries oxygenated blood from the lungs to the heart, the latter returns blood from all other tissues to the heart, either directly or through the portal system. Unlike the corresponding arteries, the systemic veins may be divided usually into a superficial and a deep set; those which generally accompany the arteries are called venae comitantes and are usually of the deep set There are commonly rather extensive connections between the two sets in any particular region

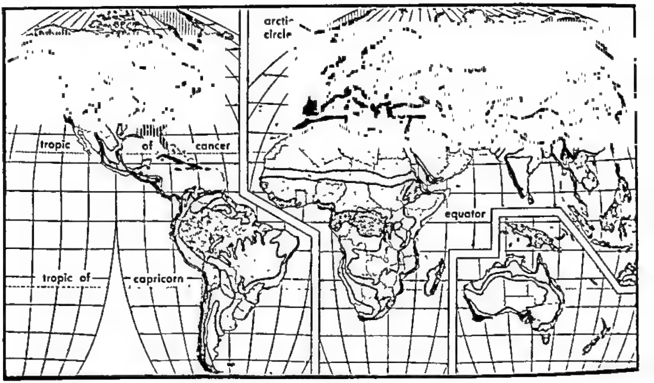
Only a few specific veins persist throughout the vertebrates, and in most cases each class or order has developed its own peculiarities of venous drainage, many of which are difficult to correlate phylogenetically The added factor of wide individual variation is also encountered

In the adult human, the veins of the head, neck, upper extremities, and upper thorax empty into the large superior vena cava. Blood from the abdomen and lower parts returns to the inferior vena cava Both of the caval vessels empty into the right atrium of the heart A few veins, such as the coronary veins, drain directly into the heart itself.

An unusual and important feature of the venous circulation in vertebrates, including man, is the presence of a portal system Most of the blood from the gastrointestinal tract, spleen, and pancreas passes into the hepatic portal vein, rather than into the inferior vena cava The portal vein opens into a rich network of sinusoidal capillaries in the liver These, in turn, merge to form veins which join together as the hepatic vein, which empties into the inferior vena cava This system is apparently a device to bring the blood from the digestive tract into intimate contact with the cells of the liver so that food storage and other liver functions may be best carried out.

Other special venous adaptations are also encountered in various species or in certain tissues of an individual See CARDIOVASCULAR SYSTEM

[E.C.ST.]



LEGEND

low-latitude forests
- tropical rain forest
- lighter tropical forest (semideciduous)
- scrub and thorn forest

middle-latitude forests
- Mediterranean scrub forest
- broadleaf and mixed broadleaf-coniferous forest
- coniferous forest

grasslands
- savanna
- prairie
- steppe (tropical and middle latitude)

deserts
- desert shrub and desert waste
- tundra
- ice caps
- undifferentiated highland

World distribution of the principal vegetative formations. *(From Climate and Man, USDA Yearbook of Agriculture, 1941)*

to the desert and grassland margins near the southern coast, and including the western part of Madagascar. Most of the western part of the Indian plateau and the northern parts of Australia bear scrub forests or savannas

Mid-latitude mixed forest areas. The humid, warmer parts of the mid-latitudes were originally covered by forests of broadleaf trees, which in areas of poorer soil were supplanted by or mixed with conifers. This association is collectively known as the mid-latitude mixed forest. In a general way it occurs both east and west of the continental dry interiors, between latitudes 20 and 40 on the east and 40 and 60 on the west side. Such forests originally occupied the eastern part of the United States and southern Canada, western Europe from Spain to Finland and extended in a narrowing belt into central Siberia and most of eastern China In all of these places as well as in Japan, Korea, and Manchuria, the original forests have been greatly altered or removed entirely in the processes of agricultural land use. The Southern Hemisphere distributions of the mid-latitude mixed forests consist of the narrow coastal strip of south Chile, southeastern Australia, and New Zealand. In these last two areas the broadleaf trees are evergreen in contrast to their general deciduous nature elsewhere.

Mediterranean zones. Areas of winter rain and summer drought on the southwestern parts of the mid-latitude continents present Mediterranean woodland shrub and scrub forest as characteristic markers of that climatic type. This vegetation occurs in southern California, middle Chile, a small tip of South Africa, a band around almost all of the Mediterranean Sea and in southeastern and southwestern Australia

Mid-latitude grasslands. Between the mid latitude forests and the deserts is the area of the prairie and steppe grasses. These occur where the amount of rainfall is enough to cause a zone of soil moisture which will support grass but is insufficient to continue this moisture zone deep enough to supply tree roots Where the moisture zone is less than 2 ft deep only the short, steppe grasses can live, where it extends lower, the deeper-rooted prairie grasses are supported. In an intermediate zone these two types merge to form the mixed grass prairies The boundaries between these three grassland types depend upon the amount and distribution of the rainfall, the temperature, and the permeability of the soil; they vary somewhat from rainy to dry climatic periods. Except in North America they are imperfectly known and mapped The mid-latitude grasslands ... tween the

Angular displacement. Figure 3 represents a body rotating with circular motion about an axis through O perpendicular to the figure. Line OP_1 is the position of some radius in the body at a time t_1, with θ_1 being the angular displacement from a reference line. Line OP_2 is the position of the same radius at a later time t_2, with the angular displacement θ_2. Angular displacement may be measured in degrees, radians, or revolutions.

Angular speed. From Fig. 3, it is seen that the body has rotated through the angle $\Delta\theta = \theta_2 - \theta_1$ in the time $\Delta t = t_2 - t_1$. The average angular speed $\bar{\omega}$ is defined by $\bar{\omega} = \Delta\theta/\Delta t$, the instantaneous angular speed ω being $\omega = d\theta/dt$. Although it is customary in most scientific work to express angular speed in radians per second, it is common in engineering practice to use the units of revolutions per minute (rpm) or revolutions per second (rps).

Tangential velocity. When a particle rotates in a circular path through an angular distance $\Delta\theta$ in a time Δt, as in Fig. 3, it traverses a linear distance Δs. The average linear speed $\bar{v}$ is given by

$$\bar{v} = \frac{\Delta s}{\Delta t} = \frac{R\,\Delta\theta}{\Delta t} = \bar{\omega}R$$

since $\Delta s = R\,\Delta\theta$. Similarly, the instantaneous speed v is given by $v = \omega R$. The direction of this instantaneous speed is tangential to the circular path at the point in question. Any vector v drawn in this direction represents the tangential velocity.

Combined velocities. A body may have combined linear and angular motions, as is the case when the wheel of a moving automobile rolls along the ground with an angular velocity about its axle which moves with a linear velocity parallel to the pavement. In this case, a point on the tread of the tire describes a curved path called a cycloid. If a circular body rolls on the surface of a sphere, a point on the periphery of the rotating body describes a curve called an epicycloid. *See* CYCLOID, EPICYCLOID. [C.P.H.]

Bibliography. H. Goldstein, *Classical Mechanics*, 1950.

Velocity analysis

A technique in machine design for the determination of the velocities of parts of a mechanical device. Velocities of various points within a mechanism must be ascertained to determine output velocity for a given input velocity, or vice versa, or to determine acceleration forces acting in the mechanism. An elementary approach is to analyze graphically simple component linkages of the mechanism that are in a single plane. Two representative components will be considered here: (1) four-bar linkage, and (2) cam and follower. For wheels in rolling contact on fixed ...

ds ...
ity ...
ma ...
ari... ...city

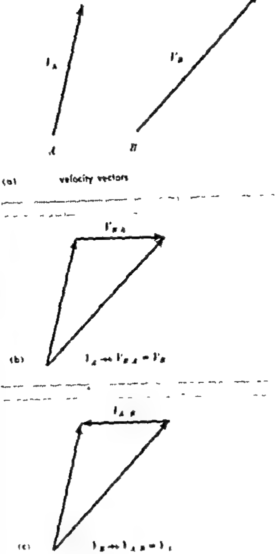


Fig. 1 (a, b, c) Relative velocity.

Angular velocity of a member V/r, where r is radius, implies the existence of an instantaneous center of rotation.

Relative velocity. All velocities are expressed relative to a reference velocity. Point A (Fig. 1a) has velocity represented by vector V_A. Point B has velocity represented by vector V_B. The relative velocity of B with respect to A, designated $V_{B/A}$, is determined by adding vectors (Fig. 1b). Relative velocity of A with respect to B is determined similarly (Fig. 1c). By convention, velocity relative to a fixed member of a linkage, which member is fixed relative to the earth, is termed absolute velocity.

Two fundamental methods of velocity analysis are outlined here: the velocity polygon and the method of instant centers.

Velocity polygon (relative velocities). Given ω_1 (Fig. 2a), find V_C. The first step is to find $V_B = r\omega_1 = \overline{AB}\omega_1$. With V_B now known, draw vector V_B from arbitrary pole O (Fig. 2b) with direction perpendicular to $\overline{AB}$. Because B and C are points

Velocity

The time rate of change of position of a body in a particular direction. Linear velocity is velocity along a straight line, and its magnitude is commonly measured in such units as meters per second (m/sec), feet per second (ft/sec), and miles per hour (mph). Since both a magnitude and a direction are implied in a measurement of velocity, velocity is a directed or vector quantity, and to specify a velocity completely, the direction must always be given. The magnitude only is called the speed (*see* SPEED). The angular velocity of a body undergoing circular motion is a vector quantity, such as $\mathbf{v}_1$ or $\mathbf{v}_2$ in Fig. 3, whose magnitude is the angular speed and whose direction is along the axis about which the rotation takes place.

Linear velocity. A body need not move in a straight line path to possess linear velocity. When a body is constrained to move along a curved path, it possesses at any point an instantaneous linear velocity in the direction of the tangent to the curve at that point. The average value of the linear velocity is defined as the ratio of the displacement to the elapsed time interval during which the displacement took place. The displacement of a body from an initial position s_0 to a final position s_f after time t is equal to $s_f - s_0$. The corresponding time interval is $t_f - t_0$. The magnitude of the velocity is then

$$v = \frac{\text{displacement}}{\text{elapsed time}} = \frac{s_f - s_0}{t_f - t_0} = \frac{s}{t}$$

where s stands for displacement and t stands for the corresponding elapsed time.

The magnitude of the instantaneous velocity of a body is the limiting value of the foregoing ratio as the interval approaches zero. In symbols this is

$$\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt}$$

where Δs and Δt are infinitesimals and ds/dt is the notation of calculus representing the time rate of change of displacement (Fig. 1).

The velocity of a body, like its position, can only be specified relative to a particular frame of reference. Consequently, all velocities are relative. *See* RELATIVE MOTION. [R.D.BU.]

Angular velocity. The representation of angular velocity ω as a vector is shown in Fig. 2. If a body rotates simultaneously about two or more axes, the resultant angular velocity is the vector sum of the individual angular velocities. Thus, if a body rotates about an x axis with an angular velocity ω_x, and simultaneously about a y axis with an angular velocity ω_y, the resultant angular velocity ω is

$$\omega = \omega_x + \omega_y$$

It should be emphasized that whereas angular velocities are commutative in addition, that is, they may be added in any order, angular displacements are not commutative. *See* ROTATIONAL MOTION.

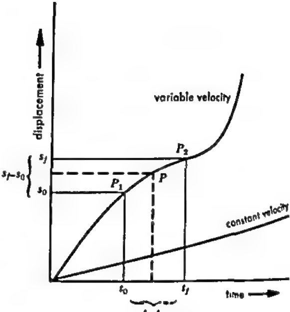


Fig. 1. The average velocity from P_1 to P_2 $(s_f - s_0)/(t_f - t_0)$. The instantaneous velocity at poi P is the limit of the ratio representing the avera velocity as the interval approaches zero.

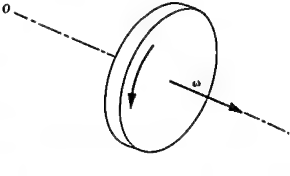


Fig. 2. Angular velocity shown as a vector.

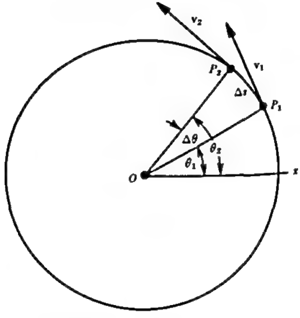


Fig. 3. Illustrating angular dis ... angul speed, and tangential velocity

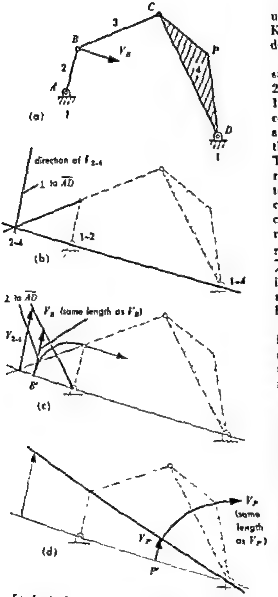


Fig 6 (a, b, c, d) Velocity analysis by means of instant centers

instant center, or centro. The instant center may be a pin joint, as A (Fig 4); it may be a finite point located physically on neither link; or, when parallel relative displacement of two links occurs, the instant center will be located at infinite distance from both links.

By definition, an instant center is a point of common velocity on any two links of a mechanism. It is a coincident point on two links at which the relative velocity of the links is zero. It is also, as stated above, the point about which one link rotates with respect to the other.

The number of instant centers for a linkage is equal to the number of possible combinations of n links, taken two at a time. This is $n(n-1)/2$. A four-bar linkage has 6 instant centers; a six-bar linkage has 15 instant centers.

Instant centers can be located systematically by using (1) Kennedy's theorem (Alexander B. W. Kennedy, 1847–1928) and (2) a finding diagram, described below.

Kennedy contemplated any three links in the same plane (Fig. 4). If the links are designated 1, 2, and 3, the possible instant centers may be called 1-2, 2-3, and 3-1. Instant center 1-2 is the point of common velocity on links 1 and 2, 2-3 on links 2 and 3, and so forth. The theorem states that the three instant centers must lie on a straight line. Thus, in Fig 4, the instant centers 1-2 and 3-1 are readily located at pin joints A and B. Instant center 2-3 must lie on line AB. At C, for example, the coincident points C on 2 (C2) and C on 3 (C3) cannot have identical velocity vectors, because V_{C2} must be perpendicular to the line AC and V_{C3} must be perpendicular to line $\overline{BC}$. Only along line $\overline{AB}$ can the vectors V_{C2} and V_{C3} coincide. Depending on relative angular displacement of links 2 and 3, instant center 2-3 may lie to the left of A, between A and B, or to the right of B.

All instant centers for a four-bar linkage are indicated in the finding diagram (Fig. 5). This diagram is useful if the dotted lines are made solid as instant centers are found. Each triangle in the diagram refers to a Kennedy line of three centers. For example, triangle $\overline{124}$ in the finding diagram represents the line of centers of links 1, 2, and 4, which line contains instant centers 1-2, 2-4, and 4-1. Line $\overline{24}$ in the finding diagram is common to triangles $\overline{124}$ and $\overline{234}$. That is, instant center 2-4 must lie at the intersection of the lines of centers represented by the two triangles. The utility of this diagram becomes apparent when linkages of five or more links are analyzed. Instant centers need be found only for links whose velocities must be determined.

Given V_B (Fig. 6a), find V_P, a point on link 4 by instant centers. In this problem, a velocity on

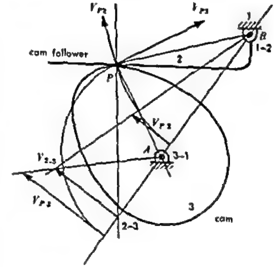


Fig 7 Instant center for links in sliding contact

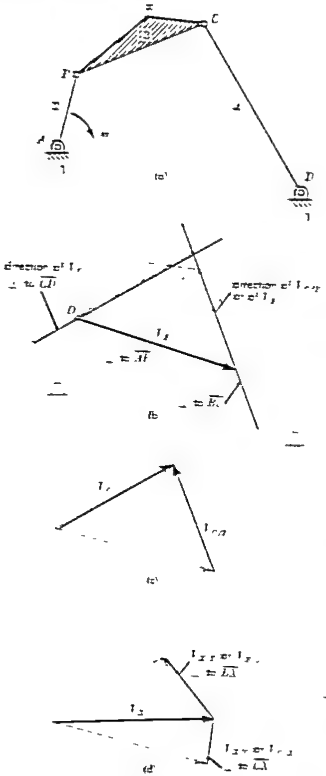


Fig. 2. (a, b, c, d) Velocity analysis by means of a velocity polygon.

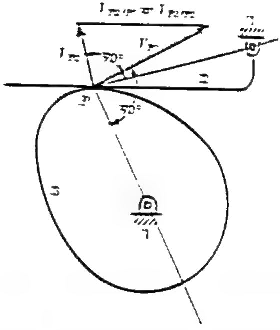


Fig. 3. Velocity polygon for links in sliding contact.

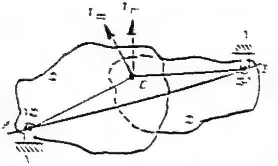


Fig. 4. Demonstration of Kennedy's theorem.

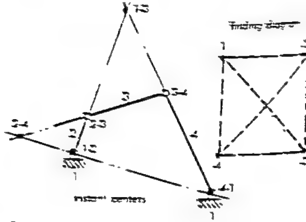


Fig. 5. Use of finding diagram for instant centers.

of the same rigid link, the direction of V_{CB} must be perpendicular to a line connecting B and C. If this were not so, link 3 would shrink or stretch. Therefore, draw a line through the end of vector V_B perpendicular to $\overline{BC}$. This establishes the direction of a vector V_{CB} or V_{BC}. Next, the direction of V_C must be perpendicular to $\overline{CD}$, by observation. Thus, draw this line of direction through pole O, from which pole absolute velocity vectors (velocity relative to fixed member 1) must radiate. Finally, by vector addition $V_C = V_B + V_{CB}$ (Fig. 2c). From these vectors, the velocity of an intermediate point E can be found (Fig. 2d).

For links in sliding contact (Fig. 3), relative velocities of coincident points on adjacent links must have directions perpendicular to a common tangent of the two links. Thus, if velocity V_{P3} of point P on link 3 is known, velocity V_{P2} of point P on link 2 can be found as shown.

Instant centers. Any two links of a kinematic linkage must, at any instant, rotate relative to each other about a common point [illegible] lled an

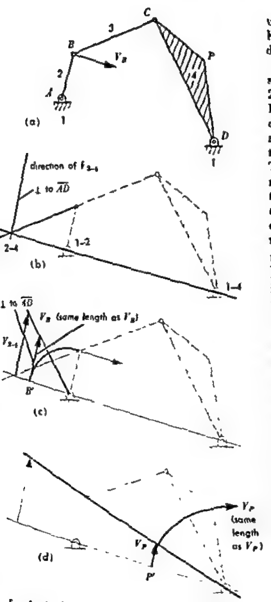


Fig 6 (a, b, c, d) Velocity analysis by means of instant centers

instant center, or centro. The instant center may be a pin joint, as A (Fig 4), it may be a finite point located physically on neither link, or when parallel relative displacement of two links occurs, the instant center will be located at infinite distance from both links

By definition, an instant center is a point of common velocity on any two links of a mechanism It is a coincident point on two links at which the relative velocity of the links is zero It is also, as stated above, the point about which one link rotates with respect to the other

The number of instant centers for a linkage is equal to the number of possible combinations of n links, taken two at a time This is $n(n-1)/2$ A four bar linkage has 6 instant centers; a six bar linkage has 15 instant centers

Instant centers can be located systematically by using (1) Kennedy's theorem (Alexander B W. Kennedy, 1847–1928) and (2) a finding diagram, described below.

Kennedy contemplated any three links in the same plane (Fig. 4). If the links are designated 1, 2, and 3, the possible instant centers may be called 1-2, 2-3, and 3-1. Instant center 1-2 is the point of common velocity on links 1 and 2, 2-3 on links 2 and 3, and so forth. The theorem states that the three instant centers must lie on a straight line. Thus, in Fig. 4, the instant centers 1-2 and 3-1 are readily located at pin joints A and B. Instant center 2-3 must lie on line AB. At C, for example, the coincident points C on 2 (C2) and C on 3 (C3) cannot have identical velocity vectors, because V_{C2} must be perpendicular to the line AC and V_{C3} must be perpendicular to line $\overline{BC}$. Only along line $\overline{AB}$ can the vectors V_{C2} and V_{C3} coincide. Depending on relative angular displacement of links 2 and 3, instant center 2-3 may lie to the left of A, between A and B, or to the right of B

All instant centers for a four-bar linkage are indicated in the finding diagram (Fig. 5). This diagram is useful if the dotted lines are made solid as instant centers are found Each triangle in the diagram refers to a Kennedy line of three centers. For example, triangle $\overline{124}$ in the finding diagram represents the line of centers of links 1, 2, and 4, which line contains instant centers 1-2, 2-4, and 4-1. Line $\overline{24}$ in the finding diagram is common to triangles $\overline{124}$ and $\overline{234}$ That is, instant center 2-4 must lie at the intersection of the lines of centers represented by the two triangles. The utility of this diagram becomes apparent when linkages of five or more links are analyzed. Instant centers need be found only for links whose velocities must be determined.

Given V_B (Fig 6a), find V_P, a point on link 4 by instant centers In this problem, a velocity on

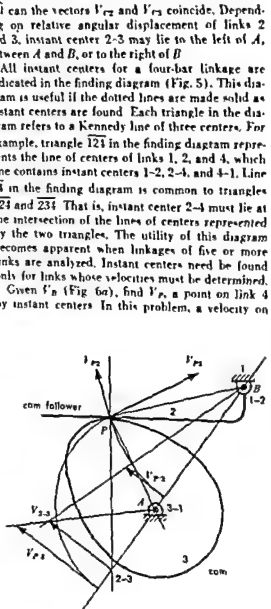
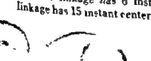


Fig. 7. Instant center for links in sliding contact



link 2 is known; the velocity of a point on link 4 is sought. Therefore, find instant center 2–4, as above (Fig. 5). Now, 2–4 is a point of common velocity on links 2 and 4. Because link 2 turns about fixed point A, the direction of $V_{2\text{-}4}$ is known, by observation, to be perpendicular to line $\overline{DA}$ extended (Fig. 6*b*).

The magnitude of $V_{2\text{-}4}$ can be found graphically by proportion. The magnitude of V'_B is equal to V_B, and direction of V'_B is parallel to $V_{2\text{-}4}$ (Fig. 6*c*). Recognizing that $V_{2\text{-}4}$ on link 2 and $V_{2\text{-}4}$ on link 4 are identical, V'_P on link 4 can be found by proportion. This velocity is then carried around fixed pivot D to yield the desired V_P (Fig. 6*d*).

In the cam and follower linkage (Fig. 7), the velocity of P on 2 can be found by using instant centers, as indicated. Notation is the same as that used in Fig 6 Instant center 2–3 must be located at the intersection of the line of three centers ($\overline{BA}$ extended) and a line perpendicular to the common tangent at point of sliding contact P. If instant center 2–3 were not along this line, sliding contact could not be maintained [E.S.F.]

Veneer

Thin sheets of wood (0 01–0 25 in. thick, depending on use) produced by rotary cutting (peeling), slicing, or (only rarely) sawing. Veneers (called face veneers) increase the usable surface of rare and expensive woods for use in facing less expensive cores in the manufacture of furniture and hardwood plywoods Veneers for construction plywood (standard veneers) are usually made of the same woods as the core (for example, Douglas-fir plywood). These standard veneers are also used in single ply in berry boxes and other containers, and in spatulas, surgical splints, and ice cream spoons *Maximum strength (technical)* veneers are the materials of choice for boat hulls, aircraft, and the like. *See* ADHESIVE; WOOD FINISHING. [C CO]

Venereal disease

A group of infectious diseases usually transmitted by sexual contact. Gonorrhea, syphilis, chancroid (soft chancre), and others less common occur in the United States.

Gonorrhea is most common and accounts for 60–75% of all venereal infections It is caused by the gonococcus bacterium, *Neisseria gonorrhoeae*, which usually produces a primary genital lesion. The organism may then spread to involve other tissues, especially joints, endocardium, meninges, and conjunctiva.

Syphilis is caused by infection with a corkscrew-shaped organism, the spirochete *Treponema pallidum* Transmission to another individual produces acquired syphilis. Transmission through the placenta to the unborn infant produces congenital syphilis.

Chancroid is an acute, localized bacterial infection by *Hemophilus ducreyi*. Granuloma inguinale, lymphogranuloma venereum, condyloma acuminata, venereal fusospirochetosis and others are also venereal diseases caused by various microorganisms *See* GONORRHEA; GRANULOMA INGUINALE; LYMPHOGRANULOMA VENEREUM; SOFT CHANCRE; SYPHILIS. [L.C.S.T.]

Ventilation

The supplying of air motion in a space by circulation or by moving air through the space. Ventilation may be produced by any combination of natural or mechanical supply and exhaust Such systems may include partial treatment such as heating, humidity control, filtering or purification and, in some cases, evaporative cooling. More complete treatment of the air is generally called air conditioning.

Natural ventilation. Natural ventilation may be provided by wind force, convection, or a combination of the two. Although largely supplanted by mechanical ventilation and air conditioning, natural ventilation still is widely used in homes, schools, and commercial and industrial buildings It is effective and economical in areas of prevailing winds and for high industrial buildings which have hot equipment which will provide the motivating convective force.

Wind-force ventilation may be provided by direct force, such as wind blowing in one side of a building and out the other side or through a monitor This is commonly known as cross ventilation Because of the friction and velocity losses incurred through building openings, only about 25–60% of the wind velocity is available for ventilation, depending upon whether the wind direction is perpendicular or oblique to the building openings. The resulting negative pressures on the downwind side of the building (if open to the outside) and the sizes and locations of the exhaust openings all materially affect the flow of air.

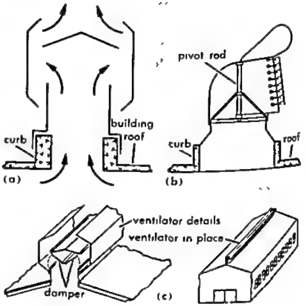


Fig 1. Roof exhausts for natural ventilation. (a) Cross section of round ventilator (b) Cross section of rotating-head ventilator (c) Continuou -ventilator

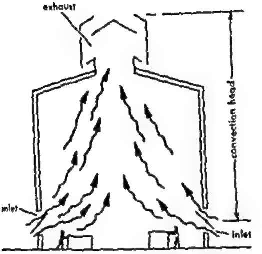


Fig 2 Building cross section illustrating motivating convective head (height) for natural ventilation

Air flow around an object creates a negative pressure on the downwind side. This principle is used to advantage in the design of a large number of building ventilators and monitors to provide basic ventilation, to assist existing convective ventilation, or to prevent backdraft down into a building. Figure 1 illustrates the round, rotating head, and continuous roof type ventilators. Because of the many complex forces involved, air-flow capacities cannot be calculated but must be determined by testing.

The force for convection ventilation is created by the difference in weight of two air columns at different temperatures, the heavy, cool column attempts to displace the hot, light column. The pressure p exerted by a fluid column varies as the height h and density u of the fluid, that is $p = hu$, expressed in consistent units. The basic equation for such flow is $V = \sqrt{2gh}$, again in consistent units

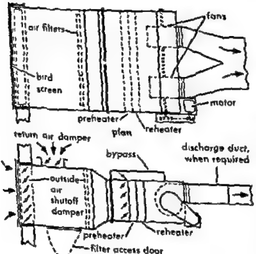


Fig 3 Factory-assembled ventilating unit

for velocity V, height h, and gravity g. From these relations and because temperature has a direct relationship to density, the following equation may be used to estimate convective flow (Fig. 2):

$$Q = 9.14A\sqrt{h(t_i - t_o)}$$

where Q = air passing through an opening by convective force, ft³/min

A = free (net) area of the opening, ft² (the inlet areas are assumed to equal the exhaust openings)

h = height from inlets to outlets, ft

t_i = average temperature of indoor air column in height h, °F

t_o = temperature of outdoor air, °F

9.14 = constant of proportionality, including value of 65% for effectiveness of openings

This equation indicates that convective velocities [illegible]

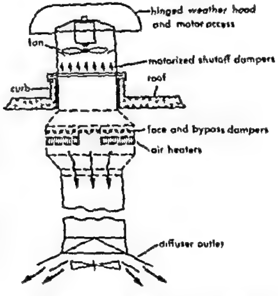


Fig 4 Mechanically powered roof air supply unit For summer ventilation portion shown solid is used Heater and dampers are for winter use. Diffuser outlet improves air distribution into ventilated space.

for convection ventilation. The best flow would be obtained by a well-shaped hole in the roof, but, because weather protection must also be provided, it is customary to exhaust [illegible]

tilation may be of the central type consisting of a central fan system with distributing ducts serving a large space or a number of spaces, or of the unitary type (Fig. 3) with little or no ductwork, serving a single space or a portion of a large space. Both types are employed for schools and for commercial and industrial applications.

Central system assemblies may be custom built,

factory prefabricated for job site assembly, or factory assembled. Ventilating units are factory assembled with rare exceptions. The assemblies consist of a fan and usually include air filters and air heaters of the finned tube type. The fans may be of the propeller, axial flow, or centrifugal type (*see* FAN). Central systems require the last two fan types because of the higher static pressures usually encountered in the distributing ducts.

The mechanically powered roof ventilator is a practical source of low-cost ventilation. For summer (nontempered) supply, such units consist of fans and weather hoods, or dampers (Fig. 4). For all-year ventilation, heaters or other equipment may be added.

Outside air connections are generally provided for all systems. Outside air is needed in controlled quantities to remove odors and to replace air exhausted from the various building spaces and equipment. The inlets are located to minimize the intake of fumes, dust, organic materials, and pollens. Because it is never possible to find completely clean air, ventilation air filters are usually provided in the system casings in all except a few industrial or summer relief applications. This prevents clogging and poor heat transfer for the air heaters and helps to reduce the pollen and dust in the occupied areas served by the systems.

Air distribution. Duct systems to distribute and disperse the air are required for all but the small unitary systems to avoid short-circuiting and to provide adequate ventilation to all parts of the space served. Such distribution permits desirable air movement in the space without undesirable draft and temperature stratification. Ventilation systems generally serve as conveyors for adding or removing heat and humidity. The distribution system must be adequate for this purpose also. The amount of air circulated is important because too small a volume requires uncomfortably high temperatures for heating or results in high building temperature for heat removal ventilation. Similar problems occur with humidity because of the limiting amounts of moisture which can be conveyed without condensation problems. Where a system is oversized, it becomes unnecessarily expensive. In addition, it may be difficult to discharge the air to the space within acceptable velocity limits.

Outlet grills and diffusers of the rectangular, square, and round type are provided to further control and distribute the supply air within the selected throw (blow length) for each outlet. Any air column acts as a pump and will entrain many times the primary volume of air. The outlets are designed and selected to obtain maximum entrainment (and mixing) because this greatly reduces air motion and temperature stratification, making possible greater comfort for the room occupants.

Exhaust ventilation systems. Exhaust ventilation is required to remove odors, fumes, dust, and heat from an enclosed occupied space. Such exhaust may be of the natural variety previously described or may be mechanical by means of roof or wall exhaust fans or mechanical exhaust systems. The mechanical systems may have minimal duct work or none at all, or may be provided with extensive ductwork which is used to collect localized hot air, gases, fumes, or dust from process operations. Where it is possible to do so, the process operations are enclosed or hooded to provide maximum collection efficiency with the minimum requirement of exhaust air.

Because of the possibilities of recirculated or external air pollution, it is customary to remove dust and fumes where practical or where required by means of ordinary ventilation filters, more efficient washers or centrifugal collectors, or chemical scrubbers.

Where dust is conveyed in the exhaust ducts, the velocities must be adequate to lift and move the dust particles. Velocities of 3000–6000 ft/min are required for this purpose. Lower velocities are tenable for fume removal, but corrosion protection must be provided by selection of duct and equipment materials. Round duct is used in most dust and fume systems because of its lower friction and better dust-handling characteristics.

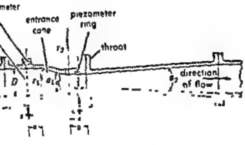
Fans for ordinary ventilation exhaust or heat removal may be similar to supply fans. Fume and dust exhaust fans must be more rugged and are frequently of the radial-blade (paddle-wheel) type for this reason. Axial flow and conventional centrifugal fans are also used in these applications. *See* COMFORT CONTROL. [J.H.CL.]

Venturi tube

A device that causes a drop in pressure as a fluid flows through it. Essentially, a venturi tube is a short straight pipe section, or throat, between two tapered sections. Local pressure varies in the vicinity of the constriction; thus, by attaching at the throat a manometer or recording instrument, the drop in pressure can be measured and the flow rate calculated from it, or, by attaching a fuel source, fuel can be drawn into the main flow stream (*see* CARBURETOR).

Proportions of the venturi tube for flow measurement, as established by its inventor Clemens Herschel, are generally as illustrated. The inlet is a short cylindrical section of the same diameter as the pipe to which it is attached. The entrance cone, having an included angle a_1, leads by an easy tangential curve into the short throat section of diameter d. A long diverging or pressure-restoring cone, having an included angle a_2, expands the fluid again to the full pipe diameter. Throat diameter ranges from one-third to three-fourths of the pipe diameter.

The pressure preceding the inlet taper is transferred through multiple openings into an annular opening, called a piezometer ring. In similar fashion, the pressure in the throat is transferred through multiple openings into another piezometer ring. A single pressure line from each ring leads to the manometer or recording meter. In some designs, the piezometer ring re replaced with



D = pipe diameter inlet and outlet
d = throat diameter as required
a = 0.25D to 0.75D for $4'' \leqq D \leqq 6''$
0.25D to 0.50D for $6'' < D \leqq 32''$
$b = d$
$c = d/2$
δ = 3/16 in. to 1/2 in. according to D
annular pressure chamber with at least 4 piezometer vents
r_2 = 3.5d to 3.75d
r_1 = 0 to 1.375D
$\alpha_1 = 21° \pm 2°$
α_2 = 5° to 15°

roportions of Herschel type venturi tube for standard nd flow measurement.

ngle pressure connections into the inlet section nd into the throat.

The principal advantage of the venturi tube is hat not more than 10–20% of the difference in ressure between the inlet and the throat is permanently lost. This is accomplished by the discharge cone gradually decelerating the flow with minimum turbulence *See* ORIFICE [R.E.SP.]

Bibliography- American Society of Mechanical Engineers, *Fluid Meters, Their Theory and Application*, 5th ed., 1959

Venus

The second planet in the solar system It is visible to the naked eye either after sunset as the evening star or before sunrise as the morning star, except for short periods near the times of its conjunctions with the Sun Its greatest elongation (greatest apparent angular distance from the Sun) is 48° Its orbit has a semimajor axis (mean distance to the Sun) of 67.7×10^6 miles. Its eccentricity of 0.007, the smallest of the main planets, causes the distance to the Sun to vary by less than 10^6 miles from aphelion to perihelion Its sidereal revolution period is 224.70 days; the mean orbital velocity, 21.9 mi/sec, and the inclination of the orbital plane to the ecliptic is 3.4° *See* PLANET

The apparent diameter of its disk varies from 10" at superior conjunction to 64" at inferior conjunction when the distance to Earth is only 26×10^6 miles, the nearest approach of any of the main planets The linear diameter, about 7,750 miles, includes the top of a cloud layer, the diameter of the solid globe must be slightly less Polar flattening is negligible The mass, about 0.815 (Earth = 1), is somewhat uncertain since, in the absence of a satellite, it must be derived from p etary perturbations. The mean density, about g/cm³, is therefore a little uncertain as is the responding value of the surface gravity, about cm/sec². By its size and mass Venus is the pl most similar to Earth

Phases. As an interior planet, Venus pre both crescent and gibbous phases, the former tween inferior conjunction and greatest ele tion, the latter between greatest elongation an perior conjunction. Because of its small dista to both Earth and Sun and its cloudy atmosp Venus appears as the brightest of the planet apparent visual magnitude varying from −3 −4.3. Maximum brightness is reached abou days before or after inferior conjunction a elongation of 39°; it is then visible to the n eye in full daylight and casts visible shado night The phase curve of Venus, giving the br ness corrected for the effect of varying distan Earth (and Sun) as a function of the phase angle Sun-Venus-Earth is shown in Fig 1, it is much closer to that of a smooth diffusing surface (perfect diffuse reflector) than the phase curves of the Moon, Mercury, or Mars and indicates the presence of a dense cloudy atmosphere. This is con firmed by the value of the visual albedo, about 0.7, the highest of the planets, indicating a reflecting power similar to that of dense terrestrial clouds (*see* ALBEDO) The color of the light reflected by Venus is slightly more yellow than direct sunlight, however, indicating that the clouds are different from those of Earth

Telescopic appearance. As viewed through a telescope, Venus usually appears as a bright, almost uniform disk or crescent, slightly shaded toward the terminator of the phase (Fig 2a). Careful examination reveals very faint, diffuse markings which vary in intensity and position from day

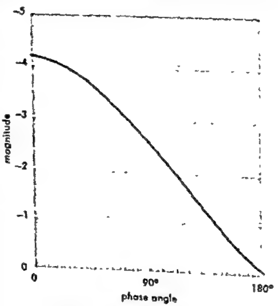


Fig 1 Phase curve of Venus

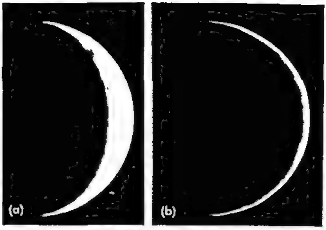


Fig. 2. Telescopic appearances of Venus. (a) Telescopic aspect. (b) Atmospheric effects

to day and are suggestive of an almost uninterrupted cloud cover Photographs in violet and ultraviolet light, however, do show strong markings, frequently arranged in rough belts, presumably parallel to the planet's equator From such evidence it may be inferred that the polar axis of the planet is not inclined more than 20° or 30° to the perpendicular to the orbital plane. Bright regions suggestive of polar caps have also been observed frequently near the cusps and support this interpretation From long series of observations a pattern of permanent or semipermanent markings can be detected in or through the constantly changing atmospheric cover This pattern has been interpreted by many astronomers, since its discovery by G. V Schiaparelli in 1889, as belonging to the surface. The relative fixity of the pattern with respect to the phase was then regarded as indicating that the planet always presents the same face to the Sun (as Mercury does), that is, that the period of rotation is equal to the period of revolution However, physical observations do not support this conclusion. Radiometric measurements indicate that the dark side of Venus is almost at the same temperature as the illuminated side, which could not be the case if the planet always had the same face toward the Sun Spectroscopic observations at the Lowell and Mt Wilson observatories indicate that the period of rotation must be much longer than 24 hours, but rotation would not have been detected if the period were longer than one or two months Hence the rotation period is still unknown as of early 1960, although it is probably longer than that of any other main planet except Mercury.

Atmosphere. The atmosphere of Venus is made directly visible near inferior conjunction by the extension of the horns of the crescent beyond 180° (Fig. 2*b*). When Venus is very close to the Sun, especially immediately before or after a transit, the whole disk appears surrounded by a faint ring of light due to scattering of sunlight by its atmosphere above the opaque cloud layer.

Spectroscopic analysis of the light reflected by Venus has revealed strong absorption bands of carbon dioxide in the red and infrared regions of the spectrum, indicating a great and variable abundance of this gas in the atmosphere. The estimated average amount present is equivalent to about 1 km at standard temperature and pressure, or almost 500 times the quantity present in Earth's atmosphere. Traces of water-vapor absorption bands have been detected in the infrared spectrum as observed from a stratospheric balloon. No oxygen is detectable from ground-based stations Bands of ionized nitrogen (N_2^+) in emission have been reported in the spectrum of the dark side of Venus; such bands might be produced by strong twilight or auroral phenomena in a nitrogen atmosphere.

The composition of the clouds is unknown Tentative identifications such as surface dust raised by convection, water-ice crystals, formaldehyde, carbon suboxide (C_3O_2) have been advanced, but none appears supported by conclusive evidence as of early 1960

Temperature. The temperature of Venus, determined by radiometric measurements in the 8-14 μ atmospheric "window" of Earth's atmosphere, is remarkably constant over the whole disk, about −38°C, the dark side is, at the most, 5 degrees colder than the illuminated side. This indicates violent convection carrying the heat from the day to the night side of the planet and a rotation sufficiently rapid to prevent carbon dioxide from freezing out of the atmosphere on the night side The measured temperature is in good agreement with the theoretical radiation temperature if the mean reflectivity (albedo) is about 0.7, as is the case in the visible spectrum. This low temperature applies to an effective radiative level in the atmosphere, and not to the surface, which must be hotter The intensity distribution among rotation lines in the carbon dioxide bands indicates in fact a higher temperature, about +12°C, which refers to a lower effective level in the atmosphere If the atmosphere of Venus is in adiabatic equilibrium the temperature gradient $dT/dh = 13°C/km$, so that the mean effective level of carbon dioxide absorption is about 3 7 km lower than the effective level of the thermal emission at 10 μ. The surface temperature derived from the thermal emission at centimeter wavelengths by means of radiotelescopes is about 580°K, or 345K° hotter than the radiometric temperature This would indicate an altitude of about

Fig 3 Optical phenomena observed during transits of Venus (a) 1874 (b) 1882

16 miles for the effective emission level of the infrared thermal radiation

Internal constitution. The internal constitution of Venus is generally supposed to be similar to that of Earth, that is, with a large iron core, but in the absence of a satellite and because of the uncertainties in mass, radius, and density the detailed internal structure cannot be derived.

Because of its general similarity to the Earth, Venus has often been considered as a potential abode of life in the solar system; however, the absence of oxygen and water vapor in its atmosphere and the high temperature of its surface seem to exclude any possibility of life

Transits. When Venus is within 1°45′ from the nodes of its orbit at inferior conjunction, a transit in front of the Sun takes place. This phenomenon is very rare and can happen only within a day or two of June 7 and December 8 when Earth crosses the line of nodes Since the first observed transit in 1639, only four have taken place—in June, 1761; June, 1769; December, 1874, and December, 1882 The next two transits will occur on June 8, 2004 and June 6, 2012 The transits of the eighteenth and nineteenth centuries were extensively observed in an effort to determine the solar parallax by a method proposed by E. Halley in 1719 However, the accuracy of the results fell far short of expectations because of optical and atmospheric phenomena New, more accurate ways of measuring the solar parallax have been developed in the meantime and it is most unlikely that the transits of the twenty first century will be observed for this application. See PARALLAX (ASTRONOMY)

Venus in transit appears as a round, black spot visible to the naked eye and crossing the solar disk in 8 hours ...

a

t

c

d ...raction of sunlight in its atmosphere above the cloud level since the lower atmosphere is opaque at grazing incidence The refraction angle, less than 1′, confirms that the phenomenon originates in the upper atmospheric layers of low density, it should not be confused with the ring of scattered light observed at inferior conjunctions outside transits

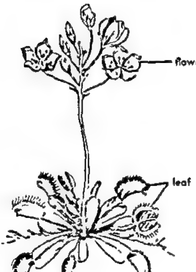
[C D V]

Bibliography: P A Moore, *The Planet Venus*, 1957

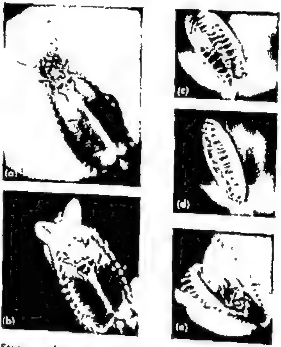
Venus' flytrap

This plant, *Dionaea muscipula*, is an insectivorous plant of North and South Carolina The two halves of a leaf blade can move as if they were hinged along the midrib, and swinging upward and inward, the two surfaces come together. Any insect alighting on a leaf triggers this sensitive motor mechanism, and is caught between the closing halves of the leaf blade In this trap, the insect is slowly digested by enzymes secreted by cells in the leaf See INSECTIVOROUS PLANTS, SARRACENIALES; SECRETORY STRUCTURES, PLANT

[P D S]



Venus' flytrap (*Dionaea muscipula*)



Stages in the capture and digestion of a fly by a leaf of Venus' flytrap (General Biological Supply House)

Verbal learning

A branch of the psychology of learning which studies the acquisition of verbal habits It deals with learning which is exhibited in everyday life in learning to spell, to read to memorize a poem, or to acquire a foreign language vocabulary. The laboratory studies from which principles of verbal learning have come, have primarily made use of single verbal units ...

sage

they

lable

... units consisting of a conso-

nant-vowel-consonant combination and which are not found in standard dictionaries, for example, "vec." The use of single verbal units greatly simplifies the laboratory control of many factors One of these factors, time of exposure of each verbal unit, is held constant through the use of a memory drum. The essential feature of this standard research device is that a series of verbal units can be presented to a learner, or subject, one at a time, but each for a constant period of time.

Types of verbal units. Verbal units are commonly formed into lists of two types In a serial list the subject is presented a group of units, one at a time, over and over in unvarying order. Learning is usually measured by instructing the subject to try to anticipate each successive unit before it is shown to him. If the units are symbolized A, B, C, D, and so forth, the subject must learn to call out B when he is shown A, to call out C when B is shown, and so on. Each unit is presented for a very few seconds on each trial, with 2 sec being a common interval. The number of units in a serial list may vary widely, although 8–12 is fairly standard. The time required to learn a list is directly related to its length. The anticipation method allows the investigator to measure the number of correct responses, as well as errors, on each trial, and the course of learning can be observed by the increase in the number of correct responses or by the decrease in errors as related to successive trials. Also, when the number of correct responses made to items at each serial position in the list is graphed, it is seen that the greatest number of correct responses is given to items in the first part of the list, the next most frequent to those near the end of the list, and the fewest to those in the middle. This bowed serial-position curve occurs under a great variety of conditions and with quite diverse kinds of materials No generally accepted theoretical explanation of the phenomenon has been given

The second major type of list, the paired-associate list, may be likened to a list of foreign words and their English equivalents. Symbolically it consists of items, A-B, C-D, E-F, etc. The first member of the pair is called the stimulus, the second, the response Each pair is presented to the subject in such a way that he first sees the stimulus alone for 2 sec and is then shown the stimulus along with the response for an equal interval of time. The subject is instructed to learn to say the response when the stimulus is presented alone Presenting the response with the stimulus after the subject attempts to give the response allows the subject to see what is correct for that stimulus The 8–12 pairs making up a paired-associate list are presented in varying orders from trial to trial so that no serial position curve develops. The progress of acquisition is observed in the same manner as in serial learning While the subject is instructed to learn to say the response when the stimulus is presented alone, it has been discovered that the subject may also learn an association in the opposite direction, that is, from the response to the stimulus. It is necessary, therefore, to think of many associations in verbal learning studies as being bidirectional.

Meaningfulness of a verbal unit. This has been defined in two ways: (1) Subjects are presented a series of units, one at a time, and are given a limited interval (2–4 sec) in which to report an association. The association value is the percentage of subjects "getting an association," and with higher percentage values are found correspondingly higher association values. (2) Subjects are presented a verbal unit for a given interval (for example, 60 sec) with instructions to write down all the different associations suggested to them by the unit. The greater the mean number of associations evoked, the higher the association value. Both methods, and certain variations on them, yield essentially the same relative association value for the same set of units. This determination of association value is preliminary to studying the relationship between meaningfulness and rate of learning Published lists of verbal units are available. These include words, nonsense syllables, and consonant syllables, which are units consisting of three consonants, and the association value for each unit. In a word list, "kitchen" would have high association value; "xylem," low. Among nonsense syllables "vaf" has a low value; "bel," high. Among consonant syllables, "qjh" has a low value; "dnt," high

Factors in rate of verbal learning. The rate of verbal learning is affected by factors like meaningfulness of a verbal unit, intralist similarity between lists of verbal units, affectivity (emotional coloring) of verbal units, transfer (influence of one task on ability to learn another task), mass versus distributed practice in learning lists, and particular differences between individuals.

Meaningfulness and rate of learning. These are directly related although the exact shape of the relationship has never been determined for the full range of meaningfulness. However, evidence indicates that the curve exemplifying this relationship will be S-shaped It is known that differences in learning at the extremes of the dimension of meaningfulness are very large For example, if a list is constructed of units of very low meaningfulness, like "rxq," and one of units of very high meaningfulness, like "cat," the latter list will be learned about three times as rapidly as the former The more analytical studies are made with paired associate lists in which the meaningfulness of the stimuli and the responses can be varied independently in different experimental conditions. When the meaningfulness of the stimuli is varied within a specified range and the meaningfulness of the responses within the same range, the rate of learning is more affected by the change in meaningfulness of the responses than by the changes in the stimuli Such findings, taken in conjunction with other results, have given some usefulness to the conception of verbal learning as a two-stage process In the first stage, the subject must learn what the responses are in the list he is learn In the

cond stage, he must associate these responses with a specific stimulus. The association value of the stimulus in a paired-associate list would primarily influence only the second stage of learning, while the association value of the response would influence both stages.

Just how differences in meaningfulness produce differences in rate of learning is a matter of some debate. A number of facts point rather conclusively to a relationship between meaningfulness and the frequency with which those verbal units have been experienced. For example, the higher the association value of a unit, the greater is the frequency with which that unit appears in written language. This holds true for nonsense syllables when viewed as parts of words and it also holds for English words. Furthermore, the experimental variation of frequency of experience with verbal units in the laboratory before these units become part of a list to be learned, appears to produce about the same effects as meaningfulness. That is, the greater the frequency, the easier the subsequent learning. It is not clear exactly what it is that must be frequent before the unit gains in meaningfulness. It is not known, for example, whether equal frequencies of seeing, hearing, or saying result in equal effects on learning. Research on study trials, interspersed with active recall attempts versus passive reading or listening, show that the former results in more learning if total learning time for the two methods is constant. Such findings suggest that the most powerful effect of frequency occurs when the subject's vocimotor system is engaged. Nevertheless, the term "frequency of experience," as related to and perhaps as a cause of meaningfulness, must be closely specified.

The frequency interpretation of meaningfulness is given indirect support by studies which use sentences having varying degrees of approximation to the structure of the language. The more closely such sentences approximate the actual structure of the language, the easier they are to learn.

Intralist similarity. This is another factor found among units of a verbal list which profoundly influences rate of learning. Similarity is varied among nonsense and consonant syllables by a nonsystematic duplication of consonants; the fewer the number of letters used in a list the higher the similarity. Similarity is varied among words by differences in similarity of meaning. The higher the intralist similarity (stopping short of identity) the slower the learning. On a common-sense level the increase in intralist similarity leads to a confusion of "what goes with what." Thus if the stimulus for one response in a paired-associate list is "unclean" and that for another response "dirty," the subject cannot use a difference in meaning as a differentiating cue. In more technical terms the inverse relationship between intralist similarity and rate of learning is said to result from interference produced by generalization among stimuli or generalization among responses (see REFLEX, CONDITIONED). In paired-associate learning, variation of intralist similarity among responses will produce less change in rate of learning than will corresponding variations among stimuli. This probably results from the fact that an increase in response similarity simplifies the response-recall stage of learning, even though it retards the association of those responses with specific stimuli. Comparable variations in intralist similarity among stimulus units result only in the interference component, that is, the interference in associating the stimuli with particular responses.

The effect of a given amount of intralist similarity, produced by repeating consonants among different nonsense or consonant syllables, is increased as the meaningfulness decreases. Although lists of high intralist similarity are difficult to learn at all levels of meaningfulness, particular arrangements of the items within the list will produce considerable benefit. For example, suppose the subject is to learn a 12 pair list in which the stimuli consist of three groups of 4 stimuli where the similarity within each group is high. If the 4 stimuli within each group are presented successively, although not in the same order on each trial, learning will be more rapid than if the 12 stimuli are presented randomly on each trial.

Affectivity of verbal units. This factor in verbal learning has also been extensively studied as a possible variable which influences the acquisition of verbal lists. Affectivity is normally defined by a rating procedure and high agreement can be reached that certain words are unpleasant, such as "vomit" and "death," while others, like "love" and "mother" are pleasant. The origin of many of these studies may be found in the desire to test Freud's theory of repression (see MEMORY). In order to study such forgetting, it is obviously necessary that the verbal units must first be learned. There is no consistent evidence that lists differing widely in the affectivity of the units are learned at different rates, if meaningfulness and intralist similarity are held constant as affectivity varies.

Transfer. This is another broad area of investigation in verbal learning. The area deals with the influence of one task on the acquisition of a subsequent one. If a preceding task facilitates the acquisition of a second one, positive transfer is said to occur; if it retards the acquisition, negative transfer is present. The most dramatic positive transfer effect is known as learning-to-learn. If a subject learns a series of lists, for example, one a day for several days, his performance on the last day will be markedly superior to that on the first day, even though all lists are known to be of equal difficulty. The effect is greater in serial than in paired-associate learning. In serial learning the subject's rate of learning may be twice as fast, for instance, on the eighth list as on the first. The processes responsible for learning-to-learn are essentially unknown. Learning-to-learn must be distinguished from warm-up. Just as a physical skill may be enhanced by warming up before performing, so also may the learning of a verbal list be

facilitated by an appropriate warm-up task. Such a task must meet the requirements that it does not involve the acquisition of specific associations which are a part of the verbal list to be learned and it must not involve learning which might be construed as a form of learning-to-learn. Such a task would be one in which the subject anticipates responses but does not form new associations. For example, the subject might be told that numbers between 1 and 9 would be presented randomly on the memory drum and he is to try to anticipate the number which will next appear. The order of the numbers will change from trial to trial so that the subject cannot, in fact, learn specific associations. A few trials on such a task are all that is necessary to produce a maximum warm-up effect (*see* PROBLEM SOLVING (PSYCHOLOGY); SENSORY LEARNING).

In studying transfer between two tasks, major emphasis has been on the effect of certain similarities between two verbal tasks. The most analytical work has been done with paired-associate lists, although both positive and negative transfer can be produced with serial lists. Similarity between tasks may be varied in the same manner as similarity within tasks. The similarity has been customarily varied between stimuli in the two lists or between the responses in the two lists. The major phenomena of transfer, if not the minor ones, resulting from variation in similarity between the stimuli and between the responses of the two lists, have been brought together fairly successfully through the use of the concepts of stimulus generalization and response generalization. Stimulus generalization is said to be a direct function of similarity of the stimuli in the two lists and response generalization a direct function of response similarity. By the theory of generalization it is assumed that, as an association develops between the stimulus and response pairs in the first list, A-B, stimuli which are similar to A will also show gain in their capacity to elicit B. The amount of gain is directly proportional to amount of generalization which in turn is directly related to degree of similarity. It may also be assumed that as the associative strength between A and B grows, responses similar to B will also have developed some associative connection with A. If pairs in the second list are constructed so as to "use" these associative connections developed as a consequence of similarity, then positive transfer in learning the second list will result. Two examples will be given. If a pair of items in the first list is "dirty—joyous," and a pair in the second list is "unclean—joyous," positive transfer in learning the second-list pair will result. Because of the similarity between the stimuli in the two lists, that is, "unclean" and "dirty," it is assumed that some associative connection was developing between the stimulus "unclean" and the response "joyous," just as the associative connection between "dirty" and "joyous" developed when the subject learned the first list. The effect is, therefore that the subject has less learning to accomplish in the second list than would be the case if he had not learned the first list; therefore positive transfer occurs. The reasoning for response generalization is the same. If a stimulus and response pair in the first list is "joyous—dirty," and a pair in the second list is "joyous—unclean," the learning of the latter pair will be facilitated.

Certain other facts complete the picture of positive transfer as a function of similarity. (1) The higher the degree of first-task learning, the greater the positive transfer. (2) Maximum positive transfer is observed on the first few trials of the second list. (3) Similarity between two responses, one in each list, has little if any effect unless there is fairly high similarity between the stimuli paired with those responses. (4) Similarity between stimuli in two lists cannot produce positive transfer unless there is fairly high similarity between the responses paired with those stimuli. If stimulus similarity is high, such as identity, and response similarity is low, the situation is one which will produce negative transfer. Thus, if a pair in the first list is "joyous—dirty," and a pair in the second list is "joyous—cold," negative transfer will be expected. As common sense would dictate, if one has learned to say "dirty" when "joyous" is shown, it will be difficult to break this habit and learn to say "cold" when "joyous" is presented. The association, "joyous—dirty," learned in the first list, must go through an extinction process and it is during this extinction period that negative transfer is apparent. In an absolute sense, the amount of negative transfer is not great in such a transfer situation. It can be maximized by constructing a second list, using the same stimuli and same responses as were in the first list but simply re-paired to form the second list.

It should be clear that an interpretation of positive transfer based on stimulus and response generalization handles only the general facts. There are less general phenomena which require additional explanatory tools. Furthermore, it is possible that theoretical tools, other than generalization, can satisfactorily handle the facts. For example, in explanation of some of the facts of transfer some investigators have used the idea of mediate association. By this conception, if a first list pair is "joyous—dirty" and a second list pair is "joyous—unclean," it is assumed that in learning the second list the subject actually has three components in his association, namely, "joyous—dirty—unclean." The word "dirty" is the mediator between "joyous" and "unclean."

Massed versus distributed practice. Massed practice is defined as a few seconds between trials, for example, 2 sec, and distributed practice as some longer interval, like 30 sec. Distributed practice will clearly enhance serial or paired associate learning only when two conditions are met: (1) The presentation rate of each item within a trial must be fairly rapid, in 2–4 sec. (2) Interference must be relatively heavy. This interference may come from intralist similarity or it may come from similarity between lists. No s[illegible]ry theoretical

explanation of the effects of distributed practice is available. When distributed practice is said to enhance learning, it must not be construed to mean that over-all time to learn a list, including the rest intervals between trials, is less than the over-all time to learn a list by massed practice. Rarely would facilitation be so great as to produce an over-all saving in time, although this could occur if the list were a long one. Normally the benefit of distributed practice is calculated merely as a reduction in number of trials required to learn to a given level of proficiency without consideration of the over-all time involved.

Individual differences. Differences in verbal learning are large even among individuals in a homogeneous group of subjects such as college students. Indeed, variations in learning attributable to individual differences are probably larger than variations which can be attributed to any other variable. It is a fact, therefore, that individuals differ widely in their speed of forming verbal associations. A part of this capacity is clearly an intelligence component, yet correlations between speed of learning verbal lists and intelligence as measured by standard tests are rarely very high. Other correlates of verbal learning of any consequence have not been discovered. In the usual anticipation procedure the subject need only anticipate responses when he thinks he knows the correct response. Obviously, some of these anticipations will be errors. Subjects show wide but consistent individual differences in their tendencies to make these anticipatory errors but the frequency with which they are made is unrelated to the speed of learning. Furthermore, if subjects are instructed in one case to maximize overt errors by guessing and in another to minimize such guessing, very little difference occurs in rate of learning. The whole area of trait differences between fast and slow verbal learners remains essentially unanalyzed.

Basic theoretical problems. Since the middle 1930s few investigators have been concerned with the conditions which must be met before any verbal associations will be formed. Rather, the concern has been with factors which cause a change in rate by which the associations are formed. Gradually, however, the work of recent years has led to a way of conceptualizing verbal learning which differs somewhat from earlier formulations. Logically the learning of any task necessitates the conception of a memory repository. Each learning trial merely represents what the learner remembers of what he learned on previous trials. The memory repository, which need not be thought of as a physical entity, even though it must be represented in the central nervous system, has at least two characteristics which may be ascribed to it. First, the evidence suggests that verbal units which have some form of similarity to each other have strong associations with each other and may be thought of as being grouped together. This may be generalized to say that any units which have associations between or among them are grouped. The second characteristic is that within a given grouping the most available response is that which has been most frequently experienced. Certain facts of verbal learning can be handled by these two principles, that is, the role of meaningfulness on the response learning phase. But the fundamental question that remains is how two units from different groupings become associated, for this is the task usually required of the subject in verbal learning. The development of this association is required before the units can be grouped together. With adults, the purity of the learning situation is decreased by the subjects' use of associational aids; these associational aids again merely beg the question of the manner in which these aids, which are already formed associations, came into being.

Although there is no general agreement on this issue, there is some reason to believe that if the confounding factors, like associational aids, could be removed, it would be found that the fundamental conditions for developing an association are frequency and contiguity of the units. Although incidental learning may be interpreted in other ways, it is quite consonant with the frequency-contiguity notion. In incidental learning, subjects are exposed to verbal units, through one guise or another, but without instructions to learn. However, learning does in fact take place and, although lesser in amount than instructed learning, appears to be a function of the same variables as instructed learning. Incidental learning takes place even though [illegible] ... are not rewarded or those that are punished. But, even in children's learning, the reward is much less obvious than it is in animal learning and even if it is obviously present may serve only as a method of giving knowledge of results, that is, to indicate the response expected of the subject. The frequency-contiguity notion is not antithetical to reward learning but may be more basic. See LEARNING THEORIES. [B.J.U.]

Vermiculite

A clay mineral constituent of clay materials. Vermiculites are similar to the montmorillonites in having an expanding structure. They differ, however, in that the expansion can take place only to a limited degree. When rapidly heated, vermiculite produces a light-weight expanded product that is widely used for thermal insulation. See CLAY, COMMERCIAL; CLAY MINERALS.

Structural properties. The structure of vermiculite consists of trioctahedral mica sheets separated by double water layers and is unbalanced by substitution of aluminum for silicon in the tetrahedral layer. The resulting charge deficiency is satisfied by exchangeable cations, usually magnesium, which occur chiefly between the mica layers.

The expansion of the vermiculite structure is restricted to two water layers. The mineral quickly expands by adsorbing two layers of water after heating to temperatures as high as 500°C. This ability to rehydrate disappears gradually above 550°C, and is completely lost at about 700°C.

The adsorbed water layers have a definite structure, but its exact nature has not been determined. As the mineral dehydrates, the structure of the water and its coordination with the oxygens of the mica structure change in stepwise fashion.

Nonstructural properties. Vermiculites show a considerable range in chemical composition. Their composition may be like that for some montmorillonites, in which case the only difference would be in the larger particle size of the vermiculites. Like the montmorillonites, the vermiculites have a high cation-exchange capacity (150 milliequivalents per 100 grams). They also absorb certain organic molecules between the mica layers, but the adsorbed organic layer is thinner and less variable than in montmorillonite.

Vermiculite is frequently listed as an alteration product of biotite mica, and it is often present along with chloritic mica as a minor constituent in ancient sediments. [F.M.W.; R.E.GR.]

Vernier

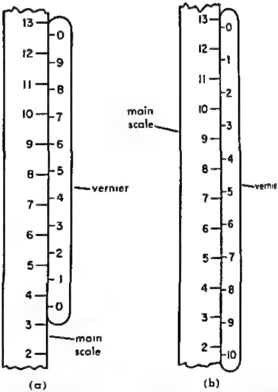
A short, auxiliary scale placed along the main instrument scale, to permit accurate fractional reading of the least main division of the main scale, invented by Pierre Vernier about 1630. The auxiliary, or vernier, scale is graduated in one or both directions from the fiducial (index) mark in numbered divisions which are fractionally shorter (in a direct vernier) or longer (in a retrograde vernier) than those on the main scale. The position of the fiducial mark (the zero mark of the vernier scale) between divisions on the main scale is indicated by the number of the graduation on the vernier scale which lines up exactly with a graduation on the main scale.

Each vernier division of the direct-reading vernier in the illustration is equal to 0.9 of the main scale divisions. The fiducial mark is located between 3 and 4 on the main scale, and since graduation 6 lines up with the main scale the exact reading is 3.6 ($3.6 + 6 \times 0.9 = 9.0$).

Such a vernier scale gives a direct reading to the nearest tenth division and permits estimation, if none of the graduations on the vernier line up exactly, to 0.03 or 0.05 divisions.

Vernier scales are used in both linear and circular measurements, as on calipers, screw micrometers, and protractors on telescope mounts and surveying instruments. The divisions are usually tenths or hundredths of units, but may be quarters, twelfths or other fractions for circular measurements. If the main scale is graduated in half degrees, the vernier divisions may be $^{11}/_{12}$ or $^{29}/_{30}$ of the main scale divisions, to permit reading to 5 or 2 minutes of arc.

The term vernier is often loosely applied to any

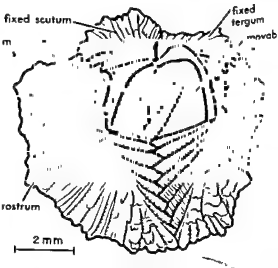


Vernier scales. (a) Direct (reading 3.6). (b) Retrogr (reading 12.7).

or more controls, such as valves or rheostats of c ferent range, that causes the least control act per unit motion or other input. *See* INSTRUMEN TION. [W.A.]

Verrucomorpha

A suborder of the Thoracica. These crustaceans sessile, asymmetrical barnacles. The plates of wall (rostrum, carina, scutum, and tergum), in single Recent genus, and the two additional pla in the fossil genus, are immovably locked togeth The unpaired scutum and tergum are movable a hinged to the wall to form a lidlike top, with o an adductor muscle. The basis is membrano caudal furca are present, and branchiae and f mentary appendages are lacking. These anim are hermaphroditic.



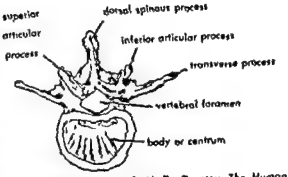
Most of the species of *Verruca* inhabit deep water. Although they have not as yet been found in certain areas of deep water, they probably have a world-wide distribution. They are small; the largest species has a diameter of about 15 mm. See CORACICA [D.P.H.]

ertebra

he cartilaginous or bony unit of the *spinal column* vertebrates. The evolutionary forerunner of the *ertebral column* was the notochord, an elastic rod ring between the spinal cord and the digestive ube. This structure still appears in the embryoogic development of every vertebrate but is replaced during early development by growth of *mesodermal tissue from* each body segment, thus forming a vertebra. In certain lower chordates, such as the Hemichorda and others, the notochord persists as the principal part of the axial, or body, skeleton. In some fishes, the vertebrae are cartilaginous throughout life, and they sometimes have an overlying bony plate. In all other vertebrates, however, the adult segments are bony units.

The vertebrae of higher forms, including man, are the result of a long series of alterations in which the original cartilaginous notochord has been replaced by specialized bone formations appropriate to the body form and habits of a species. *See* CHORDATA, HEMICHORDATA.

Human. A human lumbar vertebra exemplifies the general form and structure. It consists of a rounded, disklike body, or *centrum*, made up of hard surface bone and a somewhat spongy inner mass. A posterior neural arch forms a bony ring around the neural canal through which the spinal cord passes. Each of the neural arches in turn gives rise to three processes, the posterior spinous process and the symmetrical, paired, lateral or transverse processes. Two pairs of smaller processes arise from the roots of the neural arch. The anterior pair, the *prezygapophyses*, project backward and the posterior pair, or *postzygapophyses*, project forward. Each set has appropriate smooth surfaces, or facets, which act as articulating surfaces for successive vertebrae so that the facets of one vertebra face and join the facets of the next vertebra and so on. In addition each centrum is bound securely to its neighbors by an intervertebral disk. The center of the disk is a gelatinous material, the nucleus pulposus, a remnant of the [illegible] ...interlacing bands [illegible] lami- [illegible] ...rtilage covers the upper and lower surfaces of each vertebra, thus completing the strong, but moderately flexible, joint.



Lumbar vertebra (From R. M. De Coursey, *The Human Organism*, McGraw-Hill, 1955)

Other forms. Much variation from the previously described vertebra occurs in various classes and other subdivisions of the vertebrates; there is also marked differentiation in higher forms in regard to various regions of the spinal column in a single individual.

In fishes the presence of a ventral hemal arch in the vertebrae of the tail region distinguishes these vertebrae from those found in the trunk. Amphibians also have the two types, trunk and tail, but show the first appearance of a sacral vertebra and a single cervical vertebra, the atlas. In some amphibians, the caudal vertebrae unite to form a single long bone, the urostyle.

Cervical, trunk, sacral, and caudal regions are well defined in reptiles with legs, but are not usually recognizable in snakes or legless lizards. The first two cervical vertebrae, the atlas and axis, are quite distinct, and the centrum of the former has fused with that of the latter to form the pivot called the odontoid process which projects upward into the ring of the atlas. Most reptiles have 10 sacral vertebrae and ribs are usually present on all presacral vertebrae, except the first two. Lizards and crocodiles display a further differentiation of the trunk vertebrae into the *rib-bearing thoracic* and ribless lumbar bones.

Number and distribution of vertebrae*

	Cervical	Thoracic	Lumbar	Sacral	Caudal	Total
Man	7	12	5	5	3–5	32-34
Cat	7	13	7	3	18–25	18–53
Three-toed sloth	9	20†		6	11	36
Opossum	7	13	6	2	19–35	17-63
Chicken	16	5	3	2	9+ (pygostyle)	40+
Alligator	8	11	5	2	40+	67+
Python	(Regions not clearly differentiated)					135
Necturus	1	18†		1	23+	43+
Cyprinus		17† (body)			17	34

...ction to the *Vertebrates*, Wiley, 1933
...brae not morphologically distinguishable

In birds, all cervicals carry short ribs The thoracics, 6-10 in number, are most often fused, as are the lumbars and the sacrals. Caudal vertebrae vary in number from 9 to 12 and some of the lower bones may unite to form the pygostyle.

Mammals usually have about 35 vertebrae but the number may range from 26 to 80. All five regions are commonly present, with certain specific exceptions Seven cervicals are almost always present, 13 thoracics is the common number, but the lumbars may vary from 2 to 9. The sacrum is formed of 2-10 vertebrae and some degree of fusion is typical. The number of caudal vertebrae ranges from 2 or 3 to more than 30. In many cases they are rudimentary, as in man.

Vertebral column. The human vertebral column consists of about 33 bones. Of the 7 cervicals, the first 2 are specialized for rotation and flexion of the head. The uppermost is the atlas, an oval ring of bone with large lateral masses which articulate with the skull The second cervical, or axis, is characterized by the prominent odontoid process, articulating with the atlas, large lateral masses, and a heavy, split, spinous process. The other five cervical vertebrae are generally alike. They typically have short, often bifid, spinous processes, relatively small bodies, and openings or transverse foramina present in the transverse processes, carrying a vertebral artery, vein, and sympathetic nerves on each side of the neck region The last cervical vertebra is a transition structure with the next group of thoracic bones. Because the long, rounded spinous process is quite noticeable, this bone is given the name vertebra prominens.

The thoracic vertebrae are 12 in number in man The bodies become successively larger and each centrum carries a small pit, or facet, on each side for the articulation with the head of a rib. In addition, all but the last two thoracics also have similar facets on their transverse processes which receive the small projections, or tubercles, that are located near the heads of the ribs The spinous processes are generally long and slender, and point downward. No transverse foramina are present.

The five lumbar vertebrae are distinguished by their massive bodies, the absence of either facets or foramina, and the blunt shape of the spinous processes which point almost straight back

The five sacral vertebrae are fused in the adult to form a triangular bone, the sacrum, which is wedged between the hip bones and forms the rear wall of the pelvis The sacrum is curved backward and each of its lateral margins bears a vertical row of posterior sacral foramina through which pass certain sacral nerves and arteries.

The 3-5 coccygeal bones usually fuse and sometimes are found united to the sacrum. The coccyx, or tailbone, is the most rudimentary portion of the human spinal column.

The vertebral column of newborn infants is curved backward in the thoracic and sacral regions, presumably for better accommodation of thoracic and pelvic organs Two forward curvatures appear later, one in the cervical region when the child learns to sit up, and the second in the lumbar region when he learns to walk. These apparently represent compensation devices for maintenance of balance.

The entire vertebral column is held together by a complex system of ligaments which permit a relatively wide range of motion. Additional support and movement are obtained by the action of massive groups of back muscles, as well as supplementary muscle action of the limbs and head *See* BONE, BONE (BIOPHYSICS); EVOLUTION, ORGANIC, SKELETAL SYSTEM. [E.G.ST]

Vertebrata

A subphylum of the phylum Chordata, this group is also known as the Craniata. The vertebrates are characterized by the possession of a cranium vertebrae, brain, and visceral arches.

Vertebrates are classified by different authorities in various ways. Based on the presence or absence of jaws, two main groups can be distinguished the Agnatha and Gnathostomata. Paired appendages are lacking and no true jaws are present in the Agnatha This group includes the classes Ostracodermi, or extinct armored fishes, and the Cyclostomata, represented by the hagfishes and lampreys The Gnathostomata contains the classes Placodermi or ancient fishes, Chondrichthyes or sharks and rays, and Osteichthyes or bony fishes. These last three also comprise the superclass Pisces. These animals usually have body scales, gills, and paired fins The classes Amphibia or frogs and salamanders; Reptilia or turtles, lizards, snakes, and alligators, Aves or birds; and Mammalia or mammals constitute the superclass Tetrapoda, most species of which have paired limbs, a bony skeleton, and lungs

An alternate classification of the vertebrates is to group the Reptilia, Aves, and Mammalia in the Amniota. This scheme is based on the occurrence of a membrane, the amnion, during development All remaining groups are then included in the Anamniota, which do not possess such a structure *See* AGNATHA; AMPHIBIA; AVES; CHONDRICHTHYES; CHORDATA; GNATHOSTOMATA; MAMMALIA; OSTEICHTHYES, PISCES (ZOOLOGY), REPTILIA [C.B.C.]

Vertical take-off and landing (VTOL)

An aircraft that can rise directly up into the air and settle vertically onto the ground is a vertical take-off and landing (VTOL) craft. Such craft can operate from small airports Sketches made by Leonardo Da Vinci of a flying machine interestingly enough show a vertical take-off machine consisting of a large, fabric-covered, man powered helical screw for rising straight up into the air. The helicopter received its name from this idea The first successful flying machine, a steam engine and propeller coupled with an elongated balloon, was also the first vertical take-off aircraft with horizontal flight capabilities

The success of Orville and Wilbur Wright at Kitty Hawk in 1903 and 1904 concentrated attention on lift achieved by forward motion of large, lightly loaded wings. However, during the early years of the aviation industry some now-famous ... ch as Theodor von Kár-

... h

these early attempts were unsuccessful, the idea of flying with a vertically mounted propeller persisted, and in the late 1920s Juan de la Cierva developed the first successful aircraft of this type ... large rotor or pro-

... n

... tally in the nose of the aircraft for forward propulsion. When the aircraft moved forward, aerodynamic forces caused the vertical rotor to rotate and, thus provided the lift necessary for the craft to rise from the ground and to *maintain level flight*.

The autogiro led to the development of true powered vertical take-off, and in the late 1930s Focke-Achgelis in Germany and Sikorsky in the United States developed *successful helicopters* employing rotors directly connected to aircraft engines. Sikorsky also employed an antitorque rotor in the tail for *directional* control and to counteract the engine torque.

Whereas the powered balloon or dirigible was the first VTOL aircraft and the helicopter was the first heavier-than-air VTOL aircraft, neither could perform creditably in level flight when compared with conventional aircraft, so the search for a vertical take off and landing airplane capable of high speed flight continued. The name convertiplane has been given to a vehicle that converts to VTOL when taking off or landing. *See* CONVERTIPLANE

The inherent problem in the helicopter has been the inefficiency of propelling the propeller or rotor through the air sideways because the thrust axis of the rotor is always vertical or nearly so (see HELICOPTER). The problem as it appeared to most designers was to develop a machine in which the thrust axes could be aligned vertically for vertical take-off and landing and then aligned horizontally for normal flight. This would provide the maximum efficiency for both regimes.

Initial attempts at this were provided by the MacDonnel XV-1 and the Bell XV-3 convertiplanes developed for the U.S. Air Force in the years 1949–1959. The XV-1 was a combination helicopter autogiro employing a powered rotor for take-off and hovering and a conventional propeller for forward propulsion. The rotor was unpowered and unloaded in forward flight and the aircraft flew on a small wing. The XV-3 was of a more direct approach with tilting rotors mounted on each wing tip. These helicopter rotors were aligned vertically for VTOL and tilted to the horizontal position to act as propellers for horizontal flight.

The next step was the development of the Navy Pogo Sticks by Convair and Lockheed and the Vertijet by Ryan (Fig. 1). This time the approach

Fig. 1 Jet thrust serves to lift Ryan X-13 *tail sitter off ground, once in level flight wings provide lift.* (*Aviation Week*)

was to tilt the thrust axis by rotating the entire aircraft. These aircraft were called tail sitters; they literally sit on their tails for take-off and landing. The Vertijet was a variation of this design because it hung by a hook and cable for landing and take-off. In all these it was necessary for the pilot to have a tilting seat and to land the aircraft by looking back over his shoulder. The inability of the aircraft to operate without extensive ground equipment and the impossible loading characteristics led to abandonment of the concept.

Early in 1956 the U.S. Army funded a series of VTOL flight research aircraft (Fig. 2). These included the Vertol 76, Doak 16, Ryan Verti-plane, and Fairchild in the aircraft field, and the Chrysler, Piasecki and Aerophysics Flying Jeeps. These vehicles include the following methods of redirecting the thrust axis: (1) tilting ducted propellers, (2) tilting wing and propeller, (3) deflected slipstream,

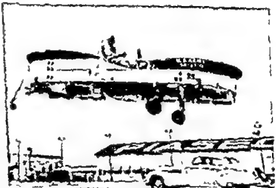


Fig. 2 An experimental *flying* jeep using ducted fans. (*Aviation Week*)

In birds, all cervicals carry short ribs. The thoracics, 6–10 in number, are most often fused, as are the lumbars and the sacrals. Caudal vertebrae vary in number from 9 to 12 and some of the lower bones may unite to form the pygostyle.

Mammals usually have about 35 vertebrae but the number may range from 26 to 80. All five regions are commonly present, with certain specific exceptions. Seven cervicals are almost always present, 13 thoracics is the common number, but the lumbars may vary from 2 to 9. The sacrum is formed of 2–10 vertebrae and some degree of fusion is typical. The number of caudal vertebrae ranges from 2 or 3 to more than 30. In many cases they are rudimentary, as in man.

Vertebral column. The human vertebral column consists of about 33 bones. Of the 7 cervicals, the first 2 are specialized for rotation and flexion of the head. The uppermost is the atlas, an oval ring of bone with large lateral masses which articulate with the skull. The second cervical, or axis, is characterized by the prominent odontoid process, articulating with the atlas, large lateral masses, and a heavy, split, spinous process. The other five cervical vertebrae are generally alike. They typically have short, often bifid, spinous processes, relatively small bodies, and openings or transverse foramina present in the transverse processes, carrying a vertebral artery, vein, and sympathetic nerves on each side of the neck region. The last cervical vertebra is a transition structure with the next group of thoracic bones. Because the long, rounded spinous process is quite noticeable, this bone is given the name vertebra prominens.

The thoracic vertebrae are 12 in number in man. The bodies become successively larger and each centrum carries a small pit, or facet, on each side for the articulation with the head of a rib. In addition, all but the last two thoracics also have similar facets on their transverse processes which receive the small projections, or tubercles, that are located near the heads of the ribs. The spinous processes are generally long and slender, and point downward. No transverse foramina are present.

The five lumbar vertebrae are distinguished by their massive bodies, the absence of either facets or foramina, and the blunt shape of the spinous processes which point almost straight back.

The five sacral vertebrae are fused in the adult to form a triangular bone, the sacrum, which is wedged between the hip bones and forms the rear wall of the pelvis. The sacrum is curved backward and each of its lateral margins bears a vertical row of posterior sacral foramina through which pass certain sacral nerves and arteries.

The 3–5 coccygeal bones usually fuse and sometimes are found united to the sacrum. The coccyx, or tailbone, is the most rudimentary portion of the human spinal column.

The vertebral column of newborn infants is curved backward in the thoracic and sacral regions, presumably for better accommodation of thoracic and pelvic organs. Two forward curvatures appear later, one in the cervical region when the child learns to sit up, and the second in the lumbar region when he learns to walk. These apparently represent compensation devices for maintenance of balance.

The entire vertebral column is held together by a complex system of ligaments which permit a relatively wide range of motion. Additional support and movement are obtained by the action of massive groups of back muscles, as well as supplementary muscle action of the limbs and head. *See* BONE; BONE (BIOPHYSICS); EVOLUTION, ORGANIC; SKELETAL SYSTEM. [E.G.S.T.]

Vertebrata

A subphylum of the phylum Chordata, this group is also known as the Craniata. The vertebrates are characterized by the possession of a cranium, vertebrae, brain, and visceral arches.

Vertebrates are classified by different authorities in various ways. Based on the presence or absence of jaws, two main groups can be distinguished, the Agnatha and Gnathostomata. Paired appendages are lacking and no true jaws are present in the Agnatha. This group includes the classes Ostracodermi, or extinct armored fishes, and the Cyclostomata, represented by the hagfishes and lampreys. The Gnathostomata contains the classes Placodermi or ancient fishes, Chondrichthyes or sharks and rays, and Osteichthyes or bony fishes. These last three also comprise the superclass Pisces. These animals usually have body scales, gills, and paired fins. The classes Amphibia or frogs and salamanders; Reptilia or turtles, lizards, snakes, and alligators; Aves or birds; and Mammalia or mammals constitute the superclass Tetrapoda, most species of which have paired limbs, a bony skeleton, and lungs.

An alternate classification of the vertebrates is to group the Reptilia, Aves, and Mammalia in the Amniota. This scheme is based on the occurrence of a membrane, the amnion, during development. All remaining groups are then included in the Anamniota, which do not possess such a structure. *See* AGNATHA; AMPHIBIA; AVES; CHONDRICHTHYES; CHORDATA; GNATHOSTOMATA; MAMMALIA; OSTEICHTHYES; PISCES (ZOOLOGY); REPTILIA. [C.B.C.]

Vertical take-off and landing (VTOL)

An aircraft that can rise directly up into the air and settle vertically onto the ground is a vertical take-off and landing (VTOL) craft. Such craft can operate from small airports. Sketches made by Leonardo Da Vinci of a flying machine interestingly enough show a vertical take-off machine consisting of a large, fabric-covered, man powered helical screw for rising straight up into the air. The helicopter received its name from this idea. The first successful flying machine, a steam engine and propeller coupled with an elongated balloon, was also the first vertical take-off aircraft with horizontal flight capabilities.

Fig 2 Crown vetch (a) In flower (b) In seed Each segment of a pod contains a seed

is slow to establish itself Diseases and insects are of minor importance in the growth of this plant. It is not known to be attacked by spittlebug (see HOMOPTERA, SPITTLE INSECT), nor has seed in storage been harmed. Seed is produced in Centre County, Pennsylvania One named variety, Penngift, was discovered in 1935 by F. V Grau

[F V G]

Bibliography: See AGRICULTURAL SCIENCE (PLANT); PLANT DISEASE.

Vibration

The term used to describe a continuing periodic change in the magnitude of a displacement with respect to a specified central reference The periodic motion may range from the simple to-and fro oscillations of a pendulum, through the more complicated vibrations of a steel plate when struck with a hammer, to the extremely complicated vibrations of large structures such as an automobile on a rough road Vibrations are also experienced by atoms, molecules, and nuclei.

A mechanical system must possess the properties of mass and stiffness or their equivalents in order to be capable of self supported *free vibration* Stiffness implies that an alteration in the normal configuration of the system will result in a restoring force tending to return it to this configuration Mass or inertia implies that the velocity imparted to the system in being restored to its normal configuration will cause it to overshoot this configuration It is in consequence of the interplay of mass and stiffness that periodic vibrations in mechanical systems are possible

Mechanical vibration. This is the term used to describe the continuing periodic motion of a solid body at any frequency When the rate of vibration of the solid body ranges between 20 and 20 000 cycles per second (cps), it may also be referred to as an acoustic vibration, for if these vibrations are transmitted to a human ear they will produce the sensation of sound (see MECHANICAL VIBRATION) The vibration of such a solid body in contact with a fluid medium such as air or water induces the molecules of the latter to vibrate in a similar fashion and thereby transmit energy in the form of an acoustic wave Finally, when such an acoustic wave impinges on a material body, it forces the latter into a similar acoustic vibration. In the case of the human ear it produces the sensation of sound. Thus, vibrations in solid and fluid bodies are essential to production, transmission, and reception of sound

One degree of freedom. Systems with one degree of freedom are those for which one space coordinate alone is sufficient to specify the system's displacement from its normal configuration. It is the simplest yet the most fundamental type of vibration system. An idealized example known as a simple oscillator consists of a point mass m fastened to one end of a massless spring and constrained to move back and forth in a line about its undisturbed position (Fig 1). Although no actual acoustic vibrator is identical with this idealized example, the actual behavior of many vibrating systems when vibrating at low frequencies is similar and may be specified by giving values of a single space coordinate They include loudspeaker cones, telephone diaphragms, microphone diaphragms, and drum membranes

When the restoring force of the spring of a simple oscillator on its mass m is directly proportional to the displacement x of the latter from its normal position, the system vibrates in a sinusoidal manner called simple harmonic motion This motion is identical with the projection of uniform circular motion on a diameter of a circle and is represented by the equation $x = A \sin 2\pi ft$ The frequency of vibration f is given by the equation $f = (1/2\pi)\sqrt{s/m}$ where s is the constant of proportionality between force and stretch or compression of the movable end of the spring The constant A represents the amplitude of the vibration, that is, the maximum displacement of the mass on either side of its rest position The magnitude of A is determined by the manner in which the motion is initially started Note that the frequency of vibration of a simple oscillator is independent of the amplitude of its vibration See HARMONIC MOTION

The variation in velocity v of the point mass of a simple oscillator is given by the equation

$$v = dx/dt = 2\pi fA \cos 2\pi ft$$

and acceleration a by the equation

$$a = d^2x/dt^2 = -4\pi^2 f^2 A \sin 2\pi ft$$

Note that for a given displacement amplitude A, the velocity amplitude $2\pi fA$ is directly proportional to the frequency, and the acceleration ampli-

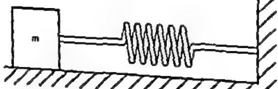


Fig 1 Simple oscillator

and (4) deflected efflux of ducted fans by use of exit vanes. These and other designs have been built and are currently being tested. The most successful methods to date are the tilting ducted fan (Doak Model 16) and the tilting wing and propeller (Vertol 76). In tests at Edwards Air Force Base, Calif., these two have shown great promise of fulfilling the joint requirements of efficient high-speed forward flight and adequate VTOL and hovering capabilities. Both of these aircraft can remain completely level throughout all flight regimes.

[N E.N.]

Bibliography: Am. Helicopter Soc., vol. 4, no. 1, 1959; N. E. Nelson, The advantages of the ducted propeller in VTOL aircraft design, *Third Annual Western Forum Am. Helicopter Soc.*, October, 1956; W. Z. Stepniewski and J. Mallen, Evaluation of VTOL systems suitable for transport application, *Aeronaut. Eng. Rev.*, September, 1957; B. J Uberti and J. B Reichert, *Design Philosophy and Test Experiences of a VTOL Aircraft*, IAS Paper 59–122, 1959

Vetch

These plants are mostly annual legumes with weak viny stems terminating in tendrils See ANNUAL PLANTS; ROSALES; STEM (BOTANY). The leaves are compound with many leaflets. *See* LEAF (BOTANY) Vetches are used mainly for green manure, cover crops, hay, and pasture Cool temperatures promote best development Seed is sown in the fall in the South and in early spring in the northern states. Inoculation of seed with nodule bacteria is necessary unless the field has grown vetch previously (*see* SOIL MICROBIOLOGY).

Hairy vetch, *Vicia villosa*, is winter hardy and is most widely grown. Smooth vetch, *V. villosa* var. *glabrescens*, and woolypod vetch, *V. dasycarpa*, are similar to hairy vetch; seeds of the three may be found in mixtures. Bird vetch, *V cracca*, is a winter-hardy perennial (*see* PERENNIAL PLANTS) Identification of vetches is difficult until pods and seeds develop (*see* PLANT TAXONOMY). Common vetch, *V. sativa*, and purple vetch, *V. bengalensis*, are less winter hardy than hairy vetch (Fig 1) Hungarian vetch, *V. pannonica*, is confined to the Pacific Northwest. Monantha vetch, *V articulata*, has fine leaves and stems but, lacking winter hardiness, is confined to warm regions. Bitter vetch, *V. ervilla*, is not grown commercially. Narrowleaf vetch, *V. angustifolia*, occurs mostly as a weed in waste places in the United States. It spreads by volunteering and makes excellent pasture. Bard vetch, *V. monantha*, is of minor importance and is grown only in irrigated valleys of the southwestern United States. Horsebeans, *V. faba*, produce coarse upright plants with large leaves and pods. *See* FRUIT (BOTANY). Some varieties are used as vegetables (*see* VEGETABLE GROWING).

Vetch seed is produced commercially along the Pacific Coast and in the southern Great Plains *See* SEED (BOTANY). Much hairy vetch seed is imported from Europe. Vetch is grown alone and with small

Fig 1. Purple vetch, *Vicia bengalensis*, in flower (USDA)

grains which give mechanical support to the weak stems of the vetch. Seed retains vitality for 5 years or longer.

Vetch diseases and insect pests. Diseases of vetch include anthracnose, gray mold, downy mildew, black stem, stem rot, root rot, rust, ovularia leaf and stem spot, and septoria scald. Root knot nematodes attack all varieties (*see* NEMATODA) Control is aided by using uninfected seed of resistant varieties on clean land

Insect pests include aphids, corn earworm, grasshoppers, cutworms, fall armyworm, weevils, and leafhoppers (*see* INSECTA). Aphids are the most troublesome and destructive (*see* HEMIPTERA).

Crown vetch. This plant, *Coronilla varia*, is a long-lived, winter-hardy perennial legume, but it is not a true vetch. It spreads by seeds and rhizomes to form a dense, weed-free, erosion resisting ground cover (*see* SOIL CONSERVATION). Its greatest use is for erosion and weed control on unmowed slopes of highways, industrial developments, and military installations. Its forage value is being studied Crown vetch has wide adaptation to variations in climate and soils. It is highly regarded for its resistance to fire. Rose, pink, and white blossoms enhance its attractiveness throughout most of the growing season (Fig. 2*a*, *b*). It thrives best on well-limed and well-drained soils, requires no mowing, and rarely needs fertilization. Plantings are made with crowns (roots) or with seed that require special bacterial in Crown vetch

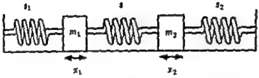


Fig 3 Simple oscillator with two degrees of freedom

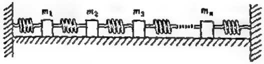


Fig 4 Example of a vibrating system with several degrees of freedom.

(Fig 4). This system has n normal modes of vibration, each of a distinct frequency.

Vibrations of elastic bodies. The primary sources of energy for producing sound waves in fluids are vibrations of elastic bodies at frequencies ranging from 20 to 20,000 cps In turn, the ultimate detection of sound waves by such devices as the human ear or a microphone requires the presence of an elastic body being forced into vibration by the impinging sound waves.

Vibrations of elastic bodies are not usually a simple harmonic motion of just one frequency but instead are of a complex nature, having many natural frequencies and modes of vibration This complexity arises from numerous factors Not all the parts of a solid body move together in phase with each other For instance, a loudspeaker diaphragm has its mass spread over a considerable surface area whose individual sectors may vibrate with different phases and amplitudes with respect to adjacent sectors Consequently, it often is necessary to give the displacement at each point on the diaphragm as a function of time in order to describe the [illegible] to specify their vibration The vibration of a solid body is influenced by interaction with its surrounding medium The natural frequencies and modes of vibration of a solid body depend upon its shape and dimensions

It is sometimes possible to confine the mode of vibration to one type having a limited number of natural frequencies The vibrations of a stretched wire are predominately transverse although it also is possible to excite weak longitudinal vibrations along the wire As the diameter of a wire increases relative to its length, the importance of the longitudinal mode of vibration increases until in a thin rod both transverse and longitudinal modes are readily excited As the diameter of a rod is further increased relative to its length, it ultimately may be regarded as a thin plate capable of vibrating in a thickness mode along its axis or in transverse modes along its surface In each of these examples it is possible to choose a method of excitation which will encourage one mode of vibration and discourage others

In any consideration of the vibration of solid bodies, characteristics of major acoustical significance include (1) the natural free vibration frequencies of the body, (2) the segmental vibration pattern for each mode of vibration, and (3) the efficiency with which the vibrations are coupled to the surrounding medium.

Vibrations of strings. The transverse vibrations of thin strings result from tension forces in the string tending to restore any displaced portion of the string to its equilibrium position. The natural frequencies f of a thin string of length l rigidly supported at its ends are given by the equation $f = nc/2l$, where n may have any integral value 1, 2, 3, . . . and c is the velocity with which transverse waves are propagated along the string. In turn, $c = \sqrt{T/m}$ where T is the tension to which the string is stretched and m is the mass per unit length of string. The fundamental, or lowest frequency mode of vibration, of a string is $c/2l$. In addition, the string is capable of vibrating at harmonic overtone frequencies of 2×, 3×, 4×, . . . this frequency (Fig. 5). The simple integral number relationship between the various frequencies of vibration of a string leads to a pleasant harmonic tonal structure which accounts for the use of vibrating strings as the primary source of sound for such musical instruments as the violin, piano, and harp. See MUSICAL INSTRUMENTS.

The relative amplitudes of the various harmonic modes of vibration of a string depend upon the particular manner in which the string is initially excited For instance, if a string is plucked at its center, the even integral (2, 4, 6, . . .) harmonic frequencies will be weak compared to the odd integral frequencies As a string freely vibrates, nodal positions of no displacement are spaced at intervals of l/n along its length For a string plucked at its center, the even harmonic frequencies are weak since all have a nodal position at the midpoint of the string

Vibrations of membranes. The transverse vibrations of stretched membranes result from tension forces in the membrane tending to restore any displaced portion of the membrane to its equilibrium position The theory of vibrations of stretched

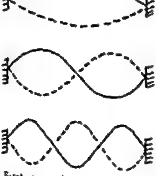


Fig 5 First, second, and third normal modes of a vibrating string

tude $4\pi^2 f^2 A$ is directly proportional to the square of the frequency. Consequently, large velocity and acceleration amplitudes of a simple oscillator are more readily obtained at high than at low frequencies.

Energy considerations. The total mechanical energy of a simple oscillator equals the sum of the kinetic energy associated with the moving mass and the potential energy stored in the distorted spring. When no frictional forces are present, these two types of energy are continuously being converted from one to the other and back again as the vibration cycle progresses. Their sum remains constant at a value of $E = sA^2/2 = 2m(\pi f A)^2$.

When frictional or other dissipative forces are present, the simple oscillator gradually changes its mechanical energy of vibration into heat energy with an attendant reduction in its amplitude of vibration. No one simple law is capable of describing the variation of frictional forces as a material body vibrates However, in many important cases the frictional force is opposite to the direction of motion and proportional to the velocity of the vibrating mass. For this type of damping force, the amplitude of vibration of the simple oscillator decreases exponentially with time in accordance with the equation $A = A_0 e^{-\alpha t}$, where α is a constant directly proportional to the frictional force and inversely proportional to the mass of the oscillator (see Fig. 2). The natural frequency of free oscillation of a damped oscillator is slightly less than that of the same oscillator system without damping. The amount by which the frequency is lowered increases with increased damping See DAMPING

External driving force A simple oscillator, or some equivalent system, is often maintained in a condition of steady vibration by the application of a periodic external driving force After a sufficient time has elapsed, any natural free-oscillation frequency of the oscillator dies out and it vibrates solely at the frequency of the impressed driving force When the applied driving force is sinusoidal, as represented, for example, by the function $F(\cos 2\pi f t)$, the forced vibration ultimately reaches a steady-state amplitude of $A = F/2\pi f Z$ where

$$Z = [R^2 + (2\pi f m - s/2\pi f)^2]^{1/2}$$

is known as the mechanical impedance of the oscillator. From this equation it is apparent that the driven oscillator will have a maximum amplitude when the forcing frequency is near $(1/2\pi)\sqrt{s/m}$, the free oscillation frequency of the oscillator. See FORCED OSCILLATION.

Nonlinear systems. When the amplitude of vibration of any real oscillator system becomes large, the elastic restoring force no longer is proportional to the displacement x. The term s relating force and stretch in the spring is not independent of x and may either increase or decrease as x increases. In either case, the motion no longer is sinusoidal and the frequency is not independent of the displacement amplitude A. The frequency becomes higher when s increases with increasing x and ... decreasing s as x increases. The latter

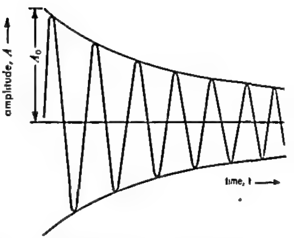


Fig. 2 Damped vibration of a simple oscillator

behavior is characteristic of the large amplitude oscillations of a pendulum bob See PENDULUM

Two degrees of freedom. When two simple vibrating systems are interconnected by a flexible connection, the combined system has two degrees of freedom Consider the simple oscillator of mass m_1 and spring s_1 connected to a second oscillator of mass m_2 and spring s_2 by means of the spring s (Fig 3) The motion of m_1 is completely described by the displacement x_1 and that of m_2 by x_2. The system is said to have two degrees of freedom since two independent space coordinates are required to specify the motion.

A system having two degrees of freedom has two normal modes of vibration of respective frequencies f' and f'' Both of these frequencies differ from the respective natural frequencies f_1 and f_2 of the individual uncoupled oscillators The larger the force constant s of the coupling spring, the larger becomes the frequency difference $(f' - f'')$ relative to $(f_1 - f_2)$, that is, f' increases relative to f_1, and f'' decreases relative to f_2.

Transverse vibrations are defined as those which occur when the vibrations of the medium in question are perpendicular to the direction of propagation of the exciting wave; longitudinal vibrations occur when the vibrations of the medium are lengthwise, along the direction of propagation of the wave An example of a vibrating system with two degrees of freedom is given by the transverse vibrations of two masses m_1 and m_2 fastened at intermediate points on a tightly stretched wire rigidly supported at each end. Here, the higher frequency mode f' corresponds to a method of vibration in which the individual motions of the two masses are oppositely directed at all times. The lower frequency mode f'' corresponds to one in which the two masses move together in phase, that is, in the same transverse direction at all times

Several degrees of freedom. A vibrating system is said to have several degrees of freedom if many space coordinates are required to describe its motion One example is n masses $m_1, m_2, \ldots, m_n$ constrained to move in a line and interconnected by $(n - 1)$ coupling springs ... al terminal springs leading from m ... sup

a judicious choice of support position may reduce or eliminate unwanted modes of vibration

An important practical application for longitudinal vibrators is their use in sonar transducers, where both the magnetostrictive vibrations of nickel tubes and the piezoelectric vibrations of crystals are utilized. *See* TRANSDUCER, UNDERWATER.

Transverse vibrations A long rod is capable of vibrating transversely as well as longitudinally, and is often difficult to produce one motion without citing the other. When a circular rod is rigidly amped at one end and free at the other (Fig 6) s fundamental frequency is $f = 0.28\ ac/l^2$ where is the velocity of longitudinal waves in the material of the rod, a is its radius, and l its length The ertone modes of vibration have frequency ratios 6.27, 17.5, 34.4, and so on to that of the fundamental

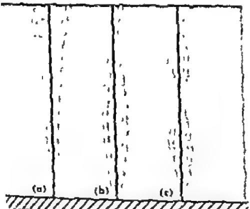


Fig 6 Transverse vibrations of a long circular rod Fundamental mode (b) First overtone mode (c) Second overtone mode

If the rod were to have a rectangular instead of a rcular cross section, its fundamental frequency ould be $f = 0.16\ tc/l^2$ where t is the lateral thickness of the rod in the direction of transverse vibration Thin rods or reeds of rectangular cross section amped at one end have numerous applications cluding use as a vibrating mouthpiece in the oodwind family of musical instruments and use as sonance vibrators in vibration frequency indicators The latter type of instrument contains a large umber of thin reeds each having a different resonance frequency When the base of the indicator is aced in contact with a vibrating object, the free d of that reed having the same natural frequency ll be set into vigorous vibration, and its visual otion can be used to indicate on an appropriate ale the numerical value of the vibration frequency.

Another important type of transverse vibration is at of a rod free at both ends The fundamental equency of such a rod of rectangular cross section $f = 0.33\ tc/l^2$, and there are overtones in ratio this frequency of 2.75, 5.4, 8.9, and so on The bars of a xylophone and similar musical instruments vibrate in this manner.

Vibration measuring equipment. Vibrations of solid bodies may result in the generation and transmission of unwanted noise. To determine the source of such noise and to devise methods for its elimination, it frequently becomes necessary to measure these vibrations.

When the vibration of a point in a material body is to be measured, a device must be provided with a sensing element which indicates either the displacements, velocities, or accelerations of the point. Such a device is usually either some mechanical system which indicates the characteristics of the vibration by means of a mechanical pointer, or a pickup capable of converting mechanical energy to electrical or some other form of energy. In conjunction with associated equipment, these pickups may be used to measure solely vibration amplitudes or to give a detailed picture of the entire vibration pattern. A vibrometer, or vibration meter, is a device indicating solely amplitude of vibration, while a vibrograph provides a complete oscillographic record of the vibration.

Vibration meter A typical electrical vibration meter consists of an electromechanical pickup, adjustable attenuator, amplifier, integrating network, and a direct reading meter capable of being calibrated to read displacement, velocity, or acceleration amplitudes Connections are also provided for oscillographic presentation, for a pair of headphones for listening to the vibration being measured, for connection to a vibration analyzer, or for connection to an electronic fre...

vi[illegible]

optical lever arms to magnify the vibratory motions *before* they are indicated or recorded These

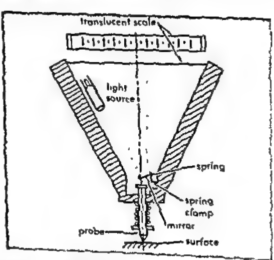


Fig 7 Mechanooptical vibrometer. The motion given to the probe by the vibrating surface is used to rock a mirror and thereby actuate an optical l...

...company)

membranes of either square or rectangular shape is primarily of mathematical interest. However, the vibrations of a circular membrane have certain physical characteristics which find practical applications. These applications include parchment membranes used for drumheads and the stretched thin steel or aluminum diaphragms of condenser microphones.

When a circular membrane is rigidly clamped along its outer radius, it will vibrate at a fundamental frequency given by the equation

$$f = (0.384/a)\sqrt{Tt/\rho}$$

where T is the tension to which the membrane is stretched, t its thickness, ρ its surface density, and a its radius When vibrating in this mode, the entire surface of the membrane moves in phase with a maximum amplitude at its center. The other possible modes of vibration are not integral multiples of f, but have frequency ratios of 2.295, 3.60, 4.90, and so forth to the fundamental. When vibrating in these latter modes, one or more nodal circles are present in which the membrane is oppositely displaced on adjacent sides of the circles The average displacement of the membrane vibrating in this manner is small, because of cancellations between the oppositely phased circular sectors This condition leads to low acoustical efficiency and limited practical significance for either the production or reception of sound For example, the response of a condenser microphone at first gradually falls off and then becomes very irregular at frequencies well above the fundamental of its diaphragm

Vibrations of plates and diaphragms. The transverse vibrations of thin plates or diaphragms result from elastic restoring forces produced when their surfaces are deformed The natural modes of vibration of such plates are involved functions of boundary shape and dimensions, whether the boundaries are free to vibrate or are rigidly clamped, and the elastic properties of its material When a thin circular plate is rigidly clamped at its rim, its fundamental frequency is given by

$$f = 0.47\frac{t}{a^2}\sqrt{\frac{Y}{\rho(1-\sigma^2)}}$$

where t is the thickness of the plate, a its radius, ρ its density, Y is Young's modulus, and σ is Poisson's ratio for the material of the plate When vibrating in this mode, the entire surface of the plate moves in phase with a maximum amplitude at its center The overtone modes of vibration have frequency ratios of 3.88, 8.70, and so on, to the fundamental When the plate is vibrating in these latter modes, one or more nodal circles are present in which the plate is oppositely displaced on adjacent sides of the circles The average displacement of the plate's surface under these conditions is small because of cancellations between the oppositely phased circular sectors As in the case of circular membranes, this condition leads to low acoustical efficiency.

The theory of transverse vibration of thin plates finds application in the design of such devices as telephone receiver diaphragms, horn-type loud speaker diaphragms, underwater sound projectors used in sonar systems, and in understanding the vibrations of wall panels, floors, automobile body panels, and hull plating of ships.

The fundamental mode of vibration of a circular plate free at its rim has a frequency some 13% lower than when rigidly clamped. The relatively involved problem of transverse vibrations of a thin flat plate becomes still further complicated when the plate has a simple curved surface, and a theoretical solution is impossible when the plate is curved into the shape of a bell

Vibrations of rods. Rod vibrations primarily consist of two simple types: (1) longitudinal vibrations along the long axis of the rod, and (2) transverse vibrations at right angles to this axis

Longitudinal vibrations. When any plane cross section in a long thin rod is displaced longitudinally relative to adjacent planes, elastic restoring forces caused by either compression or tension in the rod tend to restore the plane to its normal position These forces result in the propagation of longitudinal waves along the axis of the rod The interference of two such waves traveling in opposite directions sets up a pattern of standing waves in the rod having certain discrete frequencies The magnitude of these natural frequencies depends upon the length and material of the rod and upon the particular constraints existing at the two ends of the rod.

If the rod is either free at both ends or rigidly clamped at both ends, its fundamental frequency is given by $f = c/2l$, where l is its length and c is the velocity with which longitudinal waves are propagated in the rod When vibrating in its fundamental longitudinal mode, the length of the rod is one-half the wavelength of the waves being propagated along the axis of the rod Overtone frequencies are given by integral multiples ($n = 2, 3, 4, \ldots$) of the fundamental frequency In the case of these latter modes of vibration, nodal positions having no longitudinal displacements are spaced at intervals of $l/2n$ along the length of the rod

When the rod is rigidly clamped at one end and free at the other, its fundamental frequency is given by $f = c/4l$ Its overtone frequencies are the odd integral ($n = 3, 5, 7, \ldots$) multiples of this frequency. Numerous types of constraints may be placed on the ends of a rod, for example, stiffness of elastic springs, inertia of masses, and so forth When a rod free at one end is mass-loaded with a mass m at the opposite end, its fundamental frequency gradually decreases from $c/2l$ to $c/4l$ as the magnitude of the mass increases from zero to infinity

Whenever a given rod is vibrating longitudinally in one of its natural modes it may be supported or clamped at a nodal position without interfering with this particular mode of vibration. Since only a few modes of vibration, or in some cases only one, will have a nodal pos[illegible] a given location

automobiles when a hole in the street is hit at too fast a rate; the springs may appear to bottom out, but actually it is the shock absorber or damper. See SHOCK ABSORBER.

Some of the disadvantages of viscous damping may be overcome by using air instead of liquid as the damping medium. Air, being compressible, will add to the effective spring force with large displacements. If the air is housed within a flexible bellows, damping will be attainable horizontally as well as vertically. Such a system is illustrated in Fig. 2.

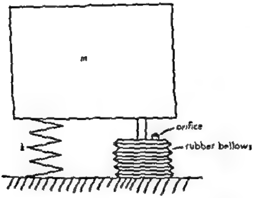


Fig. 2 System employing viscous damping with air.

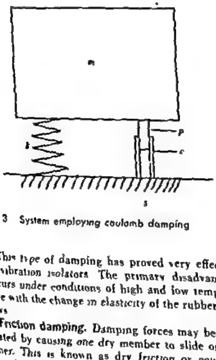


Fig. 3 System employing coulomb damping.

This type of damping has proved very effective in vibration isolators. The primary disadvantage occurs under conditions of high and low temperature with the change in elasticity of the rubber bellows.

Friction damping. Damping forces may be generated by causing one dry member to slide on another. This is known as dry friction or coulomb damping. A damper of this type is shown in Fig. 3. A pin *p* inserted in cylinder *c* and attached to body *m* is arranged to slide between two vertical spring members which are attached to the support *s*. The force exerted by the damper in opposition to the motion of the body *m* is the product of the normal force and the coefficient of friction between the pin and vertical leaf springs. The damping force is usually constant; however, if the pin is tapered, a variable force may be obtained. Friction damping is used in several commercially available isolators because it provides a simple means to control the damping forces. If necessary, independent damping systems may be provided within the same unit for vertical and horizontal motions. Some frictional dampers are effective in both vertical and horizontal directions. One such system (Fig. 4) comprises a load-carrying concave-convex spring made of a metal screen consisting of two parts attached by an eyelet. A damping-coil spring encir- [illegible] tion of the load springs results in the damping spring's being forced out over the surface of the concave springs; this creates a frictional force. Since the surface contact of the damping spring increases with displacement, the damping force is approximately linear.

Inherent damping. There are many applications where vibration dampers of an external type such as those discussed cannot be used because of space limitations, economic considerations, or the fact that the system needs very little damping. These applications make use of the inherent damping, or internal hysteresis, of such materials as rubber, felt, and cork. Vibration isolation with inherent damping is most commonly used in applications where constant motor speeds are maintained, such as in air compressors, generators, grinders, etc.

Magnetic damping. This type of damping is attainable as a result of the electric current induced in a conductor moving through a magnetic field (see INDUCTION, ELECTROMAGNETIC). The damping force can be made proportional to the velocity of the conductor moving through the field. Magnetic damping has not been used successfully in vibration isolators because its effectiveness is limited to a single direction.

Numerical values. Considerable emphasis is placed upon the numerical value for the damping force. Such a value is needed to predict the behavior of a damped system; however, difficulty is often encountered in such an analysis. The difficulty is that viscous damping, although susceptible to mathematical analysis and the establishment of a

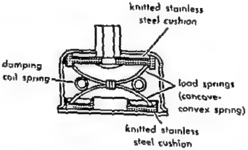


Fig. 4 Frictional damper.

devices are usually held by hand, with a projecting probe contacting the vibrating surface (see Fig. 7). Mechanical vibrometers are normally used at frequencies up to 500 cps.

Mechanical vibrographs are instruments containing a moving paper or film on which a scribing device records the amplitude of the motion being measured. In one type, shown in Fig. 8, the case

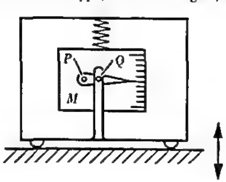


Fig. 8. Mechanical vibrograph. The scribing stylus pivots on point P, which is attached to the mass M. The case is connected to the stylus at point Q. (From C. M Harris, ed., *Handbook of Noise Control*, McGraw-Hill, 1957)

and recording paper themselves move in proportion to the motion being measured and, for sufficiently high frequencies, a mass supported on a soft spring with an attached scribing stylus serves as a stationary reference In a second type, the case and recording paper are rigidly fixed, and the vibrations are transmitted to the stylus by means of a probe in contact with the vibrating body

Resonant vibrators The simple measurement of frequency of vibrating mechanical systems is often made with an instrument containing a series of resonant mechanical vibrators These sensing elements are a series of cantilever-mounted reeds weighted at their free ends; their natural transverse frequencies are selected to cover a frequency spectrum of interest. Such instruments usually have a relatively small frequency range. By use of a series of these instruments, each covering a different frequency range, an entire range from about 20 to 500 cps may be covered.

Vibration pickups. These are electromechanical transducers capable of converting mechanical vibrations into electrical voltages. Depending upon their sensing element and output characteristics, such pickups are referred to as accelerometers, velocity pickups, or displacement pickups *See* ACCELEROMETER; VIBRATION PICKUP; *see also* SOUND; VIBRATION DAMPING; VIBRATION ISOLATION; WAVE MOTION. [L.E.K.]

Bibliography: J. P. Den Hartog, *Mechanical Vibrations*, 4th ed., 1956; L. E. Kinsler and A R Frey, *Fundamentals of Acoustics*, 1950; N. W. McLachlan, *Theory of Vibrations*, 1951.

Vibration damping

The processes and techniques used for converting the mechanical vibrational energy of solids into heat energy. While vibration damping is helpful under conditions of resonance, it may be detrimental in many instances to a system at frequencies above the resonant point. This is due to the fact that the relative motion between the base of the vibration isolator and the mounted body tends to become smaller as the isolator becomes more efficient at the higher frequencies. With damping present, the force transmitted by the elastic element is unable to overcome the damping force; this leads to a resulting increase in transmissibility. *See* DAMPING; VIBRATION; VIBRATION ISOLATION.

All metal springs which include structural members such as brackets and shelves have some damping. However, such damping is insufficient for vibration isolators and must be augmented by special damping devices.

Viscous damping. Several different types of damping devices have been developed and used successfully. Probably the most familiar is that used on automobiles, which, although known as a shock absorber, is in reality a damper, and functions as a limiter to the spring system of spring constant k. The system is shown in Fig. 1. A piston

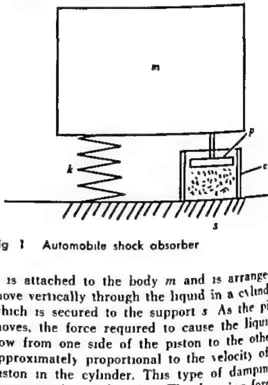


Fig 1 Automobile shock absorber

p is attached to the body m and is arranged to move vertically through the liquid in a cylinder c which is secured to the support s As the piston moves, the force required to cause the liquid to flow from one side of the piston to the other is approximately proportional to the velocity of the piston in the cylinder. This type of damping is known as viscous damping The damping force is controlled by the viscosity of the liquid and by the size of the orifice in the piston. There are several disadvantages to this type of damping; for example, it is unidirectional, it is affected by temperature changes; and because of the fact that the liquid is passed from one side of the piston to the other side through an opening, it is time conscious The opening, whether it is an orifice or the clearance between piston and cylinder, can pass only so much liquid in a given length of time. If the body to which the piston is attached is caused to displace faster than the liquid can transfer, a bottoming effect occurs. This effect is ex[illegible]ced by riders in

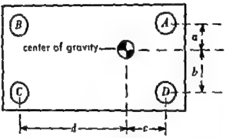


Fig 2 Determination of isolator locations

structural linkages leading to such structures, should not exceed 25% of the flexibility of the isolator. To neglect the resiliency of the support structures in providing the desired natural frequency of the system, as so many textbooks do with the assumption of rigid support, is impractical.

There are several methods for determining the load at each isolator location; one is shown in Fig. 2 where the loads are as follows

$$A = W\left(\frac{b}{a+b}\right)\left(\frac{d}{c+d}\right) \qquad B = W\left(\frac{b}{a+b}\right)\left(\frac{c}{c+d}\right)$$

$$C = W\left(\frac{a}{a+b}\right)\left(\frac{c}{c+d}\right) \qquad D = W\left(\frac{a}{a+b}\right)\left(\frac{d}{c+d}\right)$$

Here W is the total weight

The next step in vibration isolation is the positioning and arrangement of the isolators with regard to the geometry of the equipment

Location of isolators. The vibration isolators may be positioned and arranged in many different ways, all variations of three basic types each of wh...

... (2) isolators located in the plane of the center of gravity of the equipment, known as a center-of-gravity system (*see* CENTER OF GRAVITY); (3) mountings arranged four on each side in the plane of the radius of gyration, known as a double side-mounted system or radius-of gyration system (*see* RADIUS OF GYRATION)

Textbooks generally treat problems relating to the mounting of equipment with resilient elements as masses in unlimited space. Under these undefined ...

be t ...

pose ...

must be used accordingly ... and

Underneath mounting system For an underneath mounting system the most efficient arrangement for vibration isolation is one with a low natural frequency with respect to the disturbing frequency in both the vertical and horizontal axes. The most stable system is one with a h...

with respe...

case the vi...

ity may be ... with low-frequency isolators by maintaining adequate spacing of the isolators with respect to the height of the center of gravity from the mounting plane Tests show that the isolator spacing should not be less than twice the height of the center of gravity from the mounting plane. This condition is illustrated in Fig 3. The use of an underneath mounting system for equipments that exceed the condition shown in Fig. 3 should be augmented by stabilizers to provide the needed stability.

Center-of-gravity system. Locating the isolators in the plane of the center of gravity has generally been considered the ideal system because of its ability to decouple the rotational modes of vibration In actual use, however, such results are seldom achieved, because of space limitations. The primary conditions are: (1) that the isolators be located in a plane passing through the center of gravity, (2) that the distance between isolators be twice the radius of gyration of the body, and (3) that the horizontal-to-vertical stiffness of the isolators be equal. The first and last of these conditions are easy to satisfy The second condition is not always possible where a space limitation exists. This condition is generally arrived at when the height-to width ratio exceeds 2.8

Radius of-gyration system The third system that may be used when the limit of the center-of gravity system has been reached is a double side-mounted system or radius-of gyration system as illustrated in Fig. 4. Two sets of isolators are arranged on each side For optimum results the isolators should be located in the plane of the radius of gyration However, since it is often difficult to determine the exact radius, acceptable results will be obtained in most instances by assuming the body to be of uniform density Satisfactory results have been obtained with bodies having height-to-width ratios up

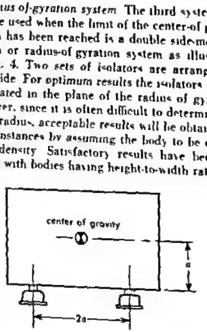


Fig 3 Underneath mounting system

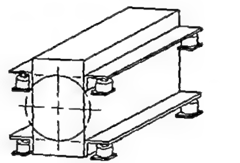


Fig 4 Radius-of-gyration system

numerical value for the damping force, is seldom encountered in pure form in actual practice. The types of damping actually used are not well adapted to mathematical analysis. The damping value may be determined by the logarithmic decrement method, but care should be exercised when these values are used in equations of pure viscous damping. This is especially true of the two most common forms of damping—friction and inherent damping. With both these types the effect of damping will vary with the amplitude of vibration. Since the damping effect is different, the natural frequency of the system will change with amplitude of vibration. For double amplitudes of vibration, such as 0.060 in., the magnification at resonance may be one value and the resonance frequency may be so many cycles per second, while with a double amplitude of 0.020 in. the magnification may be different and the resonance period may be higher. This difference must be recognized when applying formulas derived from the analysis of pure viscous damping. *See* SPRING (MECHANICAL), VIBRATION MACHINE. [K.W.J.]

Bibliography: C. E. Crede and C. M. Harris (eds.), *Handbook of Shock and Vibration Control*, 1960; J. P. Den Hartog, *Mechanical Vibrations*, 4th ed., 1956; J. N. MacDuff and J. R. Curreri, *Vibration Control*, 1958.

Vibration isolation

The isolation, in structures, of those vibrations or motions that are classified as mechanical vibration. Vibration isolation involves the control of the supporting structure, the placement and arrangement of isolators, and control of the internal constitution of the structure to be protected.

The simplest kind of mechanical vibration has the waveform of sinusoidal motion (Fig. 1). Vibrations in structures, although generally more complex in waveform, exist wherever movement takes place. Such movement may be caused, for example, by the engine in an automobile, engines or wind buffeting in aircraft, or by a punch press in a building. Delicate electronic equipment and precision instruments must normally be isolated from these motions if accurate measurements are to be obtained. *See* MECHANICAL VIBRATION, VIBRATION.

Vibration, in most cases, may be effectively isolated by placing a resilient medium, or vibration isolator, between the source of vibration and its surrounding area to reduce the magnitude of the force transmitted from a structure to its support or, alternatively, to reduce the magnitude of motion transmitted from a vibrating support to the structure. Isolating vibration at its source is commonly termed active or source isolation; isolating an instrument from its surroundings is known as passive isolation. In either case the vibration isolator is designed according to the same principles. Vibration isolators are available commercially, with published data covering their characteristics. These data include minimum and maximum rated load, natural frequency at rated load, transmissibility, damping characteristics, ultimate strength, sway space limits, etc.

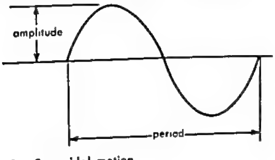


Fig. 1. Sinusoidal motion.

Natural frequency; resonance. The prime concern in the field of vibration isolation is the proper use of isolators under various load configurations with respect to their loading, the desired natural frequency, the position and location of the isolators, and the relationship to the structural response of the equipment to which they are attached. For the vibration isolator to be effective, the natural frequency should be approximately 0.4 times the frequency of the interfering source. The natural frequency is the frequency at which a freely vibrating mass system oscillates once it has been deflected. In the case where vibrations occur over a wide frequency range, such as in aircraft (5–2000+ cps), the natural frequency of the isolator is established with respect to the cruising speed. This means that when the lower frequencies are traversed, such as during aircraft takeoff, the mounted equipment will pass through a condition known as resonance. Resonance is said to exist when the natural frequency of a spring (in this case an isolator) coincides with the frequency of the excitation forces. Resonance causes magnification of the input vibration and may be harmful to the equipment if not controlled within reasonable limits. To control the vibration magnification at resonance, the resilient element within the isolator must be damped. With suitable damping the magnification factor may be held to 3 or less. *See* VIBRATION DAMPING; *see also* DAMPING; RESONANCE (ACOUSTICS AND MECHANICS).

The vibration isolator should be considered as only one part of the isolating system, the other parts being the supporting structure that lies below the isolator, and the internal structure of the equipment that is above the isolator. When isolators are selected for use where the period of resonance is critical, it should not be forgotten that the flexibility of the supporting structure and the flexibility of the isolators are in series, so that the resultant resonant frequency of the loaded system will therefore be inversely proportional to the square root of the sum of these two flexibilities. The additional flexibility of the structure will lower the natural frequency of the system and will also result in increased displacement during resonance, caused by the presence of the undamped structure. The flexibility of the supporting structure, including the

Reaction-type machine. A vibration machine in which the forces exciting the vibrations are generated by rotating or reciprocating unbalanced masses is called a reaction-type machine. The system consists of a table which is suspended from a mounting frame attached to the floor. The unbalanced masses which are attached to a shaft are part of the floating table assembly. A separate variable-speed unit, which is attached to the mass-shaft assembly of the table by a flexible coupling, drives the system

The amplitude of vibration is controlled by the degree of off-center setting of the rotating masses. Frequency is controlled by a rheostat in series with a dc motor The frequency may be continuously varied over a range of 5–60 cps. The primary advantage of this type of machine is the constant displacement provided by the rotating masses. Also, since the table is suspended, there is essentially no vibration at the floor attachment points. A reaction-type machine is pictured in Fig 2.

Electrodynamic machine. The electrodynamic vibration machine derives its vibratory force from the action of a fixed magnetic field upon a coil of wire contained in it and excited by a suitable alternating current The armature consists of those parts that are integral with the coil and the associated elements that move in the magnetic field This type works on the same principle as an electrodynamic loudspeaker The force output is proportional to the magnitude of the current Automatic displacement or acceleration control may be had over a frequency range of 5–2000+ cps

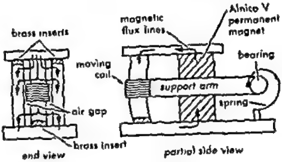
The electrodynamic vibration machine will provide ...

... is the wide frequency range it offers. Figure 3 pictures a machine complete with the control console. *See* MECHANICAL VIBRATION; VIBRATION [K W J]

Vibration pickup

An electromechanical transducer capable of converting mechanical vibrations into electrical voltages. Depending upon their sensing element and output characteristics, such pickups are referred to as accelerometers, velocity pickups, or displacement pickups

Accelerometer. This consists essentially of a mass which is seismically supported with respect to a surrounding case by means of a spring and guided to prevent motions other than those along the seismic direction of support In the operating frequency range of the accelerometer which is below its resonant frequency, the mass experiences essentially the same acceleration as does the case of the accelerometer The mass exerts a force on the spring's support which is directly proportional to the acceleration being measured This, in turn, is converted into an electrical voltage by means of stresses produced in a piezoelectric crystal For further details, *see* ACCELEROMETER



Velocity pickup. The coil swings on one end of an arm which is supported by bearings at the opposite end The case follows the motion of the structure to which it is attached (MB Manufacturing Company)

Velocity pickup. This device generates a voltage proportional to the relative velocity between two principal elements of the pickup, the two elements usually being a coil of wire and a source of magnetic field (see illustration). In a probe type of velocity pickup, the relative motion is between a coil, which is attached to one end of a probe having its other end in contact with the structure whose vibration is being measured ...

... that the voltage output of the moving coil is directly proportional to the velocity of the vibrating body.

Displacement pickup. This is a device that generates an output voltage which is directly proportional to the relative displacement between two elements of the instrument. These pickups are similar in construction and behavior to velocity pickups. The only essential difference is in the use of a frequency-weighting network required to make them direct reading Instrumentation problems in vibration measurements are generally reduced if measurements of small displacements at high frequencies are made with accelerometers or velocity pickups, since for a given displacement the acceleration varies as the square of the frequency and the velocity directly as the frequency *See* VIBRATION [L.E.K.]

Vibrator

An electromechanical device used primarily to convert direct current to alternating current but also used as a synchronous rectifier. The chopper, a closely related device is also discussed in this article.

Power vibrator. A vibrator consists of a flat reed, with a soft-iron slug mounted at the tip of the reed adjacent to a driving coil and pole piece (Fig. 1) The soft iron slug, or armature, is attracted by the pole piece of the driving coil. The longitudinal axis of the driving coil is usually parallel to the reed axis, but the pole piece is slightly offset from the neutral position of the reed so that the reed is attracted sideways toward the pole piece whenever the coil is energized.

The reed also has a set of contacts which hit stationary contacts attached to the frame. The sta-

to 5. The limitation of this system will be reached when structural rigidity of the body is such that excessive bending occurs between the upper and lower isolator locations.

The next consideration in vibration isolation is the structural rigidity of the body to be isolated. This step is of importance in that the use of incorrect structures, particularly supporting brackets for component parts, can render the other two steps useless Supporting brackets act as springs under a vibratory condition and become resonant at their natural frequency. Should resonance of the brackets coincide with the isolator resonance, damage may occur. The resonance point for the brackets and internal component should occur when the isolators are approaching their maximum efficiency so that the input vibration to the internal structure is at a low level. This level is normally reached at four times the natural frequency of the isolator.

The isolation of vibration may be accomplished in most instances by the use of commercial isolators. In some cases, such as in heavy industrial machinery where the supporting structure is concrete, less consideration need be given to the flexibility of the supporting structure. Also, where the body to be isolated is stationary, as in the case of compressors, the height-to-width ratio with respect to the isolator spacing is of little importance. *See* VIBRATION MACHINE. [K.W.J.]

Bibliography: *See* VIBRATION DAMPING

Vibration machine

A device for subjecting a system to controlled and reproducible mechanical vibration Vibration machines, commonly called shake tables, are widely used in vibration measurement and analysis. There are three types of vibration machines in general use. These are the mechanical direct-drive type, the mechanical reaction type, and the electrodynamic type. Other types, such as hydraulic excitation devices, resonant systems, piezoelectric vibration generators used for instrument calibration, and machines for testing packages, have only limited or specialized applications

Direct-drive machine. The mechanical direct-drive vibration machine (sometimes referred to as the brute-force type) is a machine in which the vibration table is forced by a positive linkage to undergo a displacement. The linkage is driven by a direct attachment to eccentrics or crankshafts The degree of eccentricity or crank radius determines the amplitude of vibration of the table. The motion of the vibration table is approximately simple harmonic (*see* HARMONIC MOTION). The frequency is varied over a range of 5–60 cps by a variable-drive unit. *Automatic cycling* is *normally* used to cycle the table to maximum frequency and return to the starting frequency. The period of cycling is usually 1–3 min.

The vibration table, by the use of several cam linkages, is usually designed to move in both vertical and horizontal directions. A machine of this type is pictured in Fig 1.

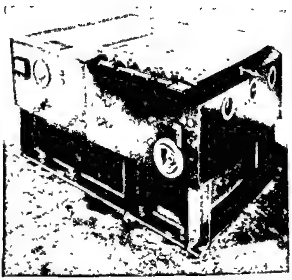


Fig 1. Direct-drive vibration machine (L.A.B Corp)

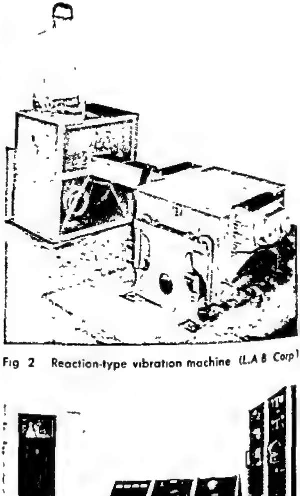


Fig 2 Reaction-type vibration machine (L.A.B Corp)

Fig 3. A complete vibration test system showing the electrodynamic shaker in the left foreground In the rear from left are the power amplifier, control console, and instrumentation cabi idyne Co)

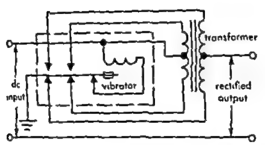


Fig 3 Vibrator circuit arranged for synchronous rectification.

former secondary A typical circuit is shown in Fig 3 The circuit is so arranged that whenever one half of the transformer primary is excited, the corresponding secondary terminal is grounded In this way the center tap of the secondary is always at the same polarity, and the output is a pulsating direct voltage This vibrator is relatively expensive, and it is generally more economical to use vacuum tubes or semiconductors to rectify the output. *See* MECHANICAL RECTIFIER

Signal chopper. To modulate low-level dc signals, an electromechanical chopper is used This device differs from a power vibrator in that the drive coil is normally excited with an alternating voltage at the desired carrier frequency, usually 60 or 400 cps A permanent magnet is incorporated in the pole structure of the coil so as to polarize the coil. The excitation of the coil alternately reinforces and reduces the ...

perman-

cutes or ...

plied voltage The driving power required is reduced to a minimum by constructing the reed to have its resonant frequency near the excitation frequency The resonant frequency is not made equal to the excitation frequency, however, because then small changes in exciting frequency would result in large phase shifts between the reed oscillations and the coil drive In a typical 60 cps chopper the reed frequency is 80 cps The contacts should not interrupt more than a few milliamperes for maximum life

The chief advantage of electromechanical choppers is their extremely low noise and drift. In this respect, they are much superior to electronic devices, and they are therefore used almost exclusively for critical applications. Choppers have been used successfully to amplify voltages as small as 10^{-9} volt (*see* CHOPPING).

A common application of choppers is in servomechanisms using an ac servo motor but having a dc error channel Here the chopper is used as a modulator to convert the dc error into an ac voltage to control the motor.

Another application of choppers is in high-gain dc amplifiers having virtually zero drift. Such amplifiers are used to amplify minute voltages, such as those generated by thermocouples They are also used in analog computers. A typical circuit is shown in Fig. 4. The dc input voltage is modulated by one contact of the chopper into an ac voltage. This voltage is amplified in a standard ac amplifier and is then demodulated by the other contact of the chopper. The signal is smoothed in a low-pass filter to reconstruct an amplified version of the original input In this application the chopper acts simultaneously as a modulator and demodulator.

[F.B.T.]

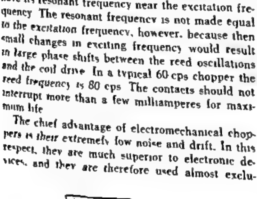


Fig 4 Chopper-stabilized dc amplifier

Vicuna

A rare animal whose fiber makes the world's most costly and most exquisite cloth, surpassing all others in fineness and beauty It is found in an almost inaccessible area of the Andes Mountains, at altitudes between 16 000–19,000 ft The vicuna, one of the wildest of animals, is less than 3 ft high and weighs 75–100 lb A single animal yields only 1/4 lb of hair; thus 40 animals are required to provide enough hair for the average coat To preserve the species, the Peruvian and Bolivian governments have placed the vicuna under their protection. Attempts to domesticate this animal have not been very successful but are still being made in Peru. The fiber of the vicuna is the softest and most delicate of the known animal fibers; yet it is strong for its weight, is resilient, and has a marked degree of elasticity and surface cohesion. It is used in costly suitings and overcoat fabrics. *See* ALPACA, CAMEL'S HAIR, CASHMERE; LLAMA, MOHAIR, WOOL; *see also* FIBER, NATURAL

[M D P.]

Video amplifier

A low-pass amplifier having a bandwidth on the order of 2–10 Mc Typical applications are in tele...

...sentially by the plate load resistance and the shunt capacitance in the circuit To extend the bandwidth of an *RC*-coupled amplifier it is necessary to overcome the effects of the capacitance. This is done by adding an *inductance*, as shown in the illustration The ind...

form a

above

employ

... The principal design requirement is to choose a value for the inductance that will extend the frequency response without introducing an undesirable hump in the gain char-

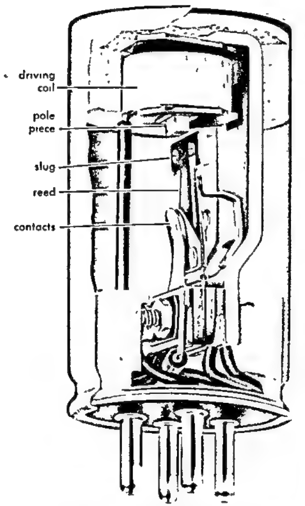


Fig. 1 Typical low-power, plug-in vibrator

tionary contact points are usually mounted on leaf springs so that they can "give" when the reed contact hits them. This minimizes contact bounce and lengthens vibrator life. The entire vibrator assembly is normally mounted inside a can and cushioned with sponge rubber. High-quality vibrators are hermetically sealed, and the terminals are brought out to a standard tube socket. Vibrators can usually withstand considerable acceleration, shock, and temperature variation; their life expectancy is comparable to that of vacuum tubes.

There are two major classes of vibrators: interrupters, in which the vibrator serves to periodically interrupt the direct current input, and synchronous vibrators, in which the vibrator periodically interrupts the input and then synchronously rectifies the ... alternating current

at rest, the two main contacts are open and the coil contact is closed. When the dc input is connected, the coil becomes energized and pulls the reed over to one side, opening the coil contact and deenergizing the coil. The reed then moves back, closing the coil contact, and the cycle repeats. The vibrations of the reed build up, and when equilibrium is reached, the reed oscillates vigorously and each of the main contacts is closed for nearly hal of the vibrator period.

The main contacts alternately switch curren from the dc supply through opposite halves of th transformer primary, setting up an alternating flu in the transformer core. Alternating voltage is then available from the transformer secondary. A resistor and capacitor connected across the transformer primary reduce sparking at the contacts.

A somewhat simpler vibrator, having only the two main contacts, is frequently used in automobile radio power supplies. In these vibrators the driving coil is connected internally between the reed and one of the stationary contacts. Other vibrator connections are the same as shown in Fig. 2. When power is first applied to the vibrator, current flows through the driving coil in series with one-half of the transformer winding. The resulting motion of the reed short-circuits the driving coil and removes the excitation. Because the driving-coil current flows through one of the transformer windings, the output of this vibrator is asymmetrical. However, because the coil current is a small fraction of the transformer load current, this is tolerable in many applications. The major advantage of this vibrator is its low cost.

Power vibrators are available for dc supply voltages ranging from 6 volts to several hundred volts. The smaller units of the type used in automobile radios can handle up to about 50 watts of power; heavy-duty vibrators having several sets of contacts have power ratings up to several hundred watts. The output frequency is usually around 120 cps; other nominal frequencies are available. The vibrator frequency is determined primarily by the resonant frequency of the reed, but is affected by contact spacing, supply voltage, and other factors.

Synchronous vibrators. Some power vibrators have an extra set of contacts either mounted on the same or on a separate (split) reed. These contacts synchronously rectify the output of the trans-

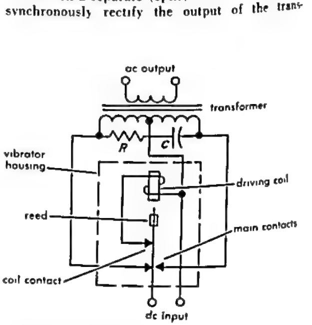


Fig. 2 Vibrator circuit for conv ting direct current to alternating current

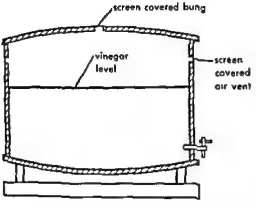


Fig 1. Orleans process barrel for making wine vinegar. (From L. A Underkofler and R J Hickey, eds, *Industrial Fermentations*, 2 vols, Chemical Publishing, 1954)

tained by the alcoholic fermentation of sugars such as syrups of glucose, black strap molasses, or refiners' molasses In the preparation of distilled vinegar, the alcohol is distilled off the fermented vegetable mashes before it is fermented by the vinegar bacteria. Nondistilled vinegars are characterized by distinctive flavors and odor. Seasoned vinegars are prepared by soaking vegetable seasoning agents, such as tarragon leaves, onions, garlic, peppers, or various spices, in vinegar

Processes. Vinegar is made by two processes, the older French or Orleans process, and the quick vinegar process

Orleans process In the Orleans process, wine or hard cider is fermented in wooden barrels or small covered vats. Air vents, with screens to exclude vinegar flies, are located in the top and upper walls of the vessel A starter culture of about equal volumes of good, nonpasteurized vinegar and fresh wine is introduced into the barrel, and the fermentation is continued until acetification is complete. About [illegible] failure to produce a good vinegar is usually due to the gradual replacement of the desirable types of vinegar bacteria by other microorganisms, including some species of vinegar bacteria themselves, which oxidize acetic acid to carbon dioxide and water The Orleans process is slow and can be operated only in small vessels The vinegar so produced is of superior flavor and aroma, due to the fact that the conditions under which the conversion of alcohol to acetic acid proceeds permit the accumulation of larger quantities of products of side reactions leading to such materials as ethyl acetate and acetoin Such vinegars are usually reserved for table use The Orleans process may be applied on a household scale

Quick vinegar process Nearly all the vinegar of commerce is now made by the quick vinegar process Here the fermentation is conducted in a vinegar generator, a vessel composed of two chambers The larger upper chamber is packed to within about 8 in of the top with solid materials having a very large surface area Beechwood shavings are preferred for distilled vinegar production and corncobs or coke for the production of wine, cider, or sugar vinegar. The packed upper chamber is separated from the lower chamber by a screen. Air is blown upward through the screen and packing material, and escapes from the top of the generator Fermenting liquids are distributed intermittently over the top of the packing, allowed to percolate through the mass, collected in the lower chamber, then pumped back to the top and recirculated until the alcohol content is reduced to one-half per cent. The vinegar is drawn off, and fresh alcoholic solution is added. This process is repeated until bacterial slimes or contamination results in reduced yield of acetic acid or poor percolation If cider, sugar, or wine vinegar is being made, a vinegar generator is cleaned and restarted twice a year, but a distilled

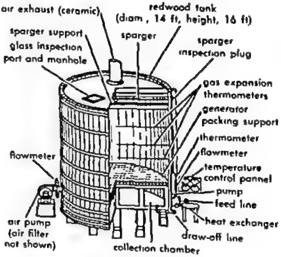


Fig 2 Recirculating-type vinegar generator. (From L A Underkofler and R. J Hickey, eds, *Industrial Fermentations*, 2 vols, Chemical Publishing, 1954)

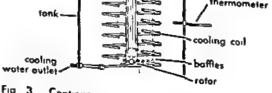
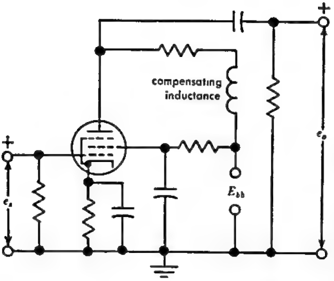


Fig 3 Continuous process generator. (From R F. Cohee and G Steffen, *Food Eng*, 31(3) 58–59, 1959)



Wiring circuit of a shunt-compensated amplifier.

acteristic for frequencies near the upper limit. An inductance can be chosen with relation to the resistance and shunt capacitance so that the bandwidth is extended but the circuit does not produce an amplification at any frequency greater than the amplification at midband frequencies It should be noted that if an amplifier composed of several stages is to have a wide bandwidth, then the bandwidth of each stage must be larger by a factor which depends upon the number of stages. This can impose severe requirements upon the design of each stage. *See* AMPLIFIER. [H F.K]

Bibliography: J D Ryder, *Engineering Electronics*, 1957.

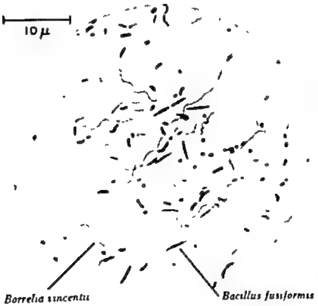
Vincent's angina

An inflammation of the tissues of the oral cavity or pharynx caused by fusiform bacilli and a spirochete *Borrelia vincenti* The underlying cause, however, is a dietary deficiency in the Vitamin B complex which favors the overgrowth of these organisms. *See* SPIROCHETE.

B. vincenti is a short, 5–10 micron (μ), actively motile, spiral organism with a variable number of shallow, irregular turns. *Fusobacterium fusiforme* is a coarse, thick, gram-negative rod, 3–10 μ in length and 0 3–0.8 μ in width These two organisms are normal inhabitants of the gums, but their numbers greatly increase in the presence of vitamin deficiencies and pyorrhea.

The commonest form of the disease, trench mouth, is characterized by redness, congestion and edema of the gums with involvement of the entire oral cavity in severe cases It is observed mainly among institutional or military groups with a deficient diet. Sporadic cases occur as a result of self-imposed dietary restrictions Severe forms of the disease may involve the tonsil, which becomes gangrenous. Rarely the mucous membrane of the bronchii is involved.

Prevention is accomplished through good oral hygiene and an adequate diet Mild cases usually respond to dietary measures alone Penicillin is used in severe cases. *See* PENICILLIN; SPIROCHAETALES. [T B T]



Spirochete and fusiform bacillus of Vincent's angina (Copyright by General Biological Supply House, Inc., Chicago)

Vinegar

The product of the incomplete oxidation to acetic acid of ethyl alcohol produced by a primary fermentation of vegetable materials. The oxidation, often called "secondary fermentation," proceeds under the influence of various kinds of so-called vinegar bacteria, species of the genera *Acetobacter* and *Acetimonas*. *See* ACHROMOBACTERACEAE; PSEUDOMONADACEAE.

The over-all chemical equations for the production of vinegar are:

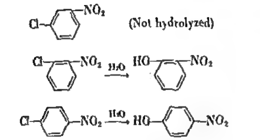
$$\text{Sugar} \xrightarrow{\text{Yeast}} 2CH_3CH_2OH + 2CO_2$$

and

$$2CH_3CH_2OH + 2O_2 \xrightarrow[\text{Bacteria}]{\text{Vinegar}} 2CH_3COOH + 2H_2O$$

Vinegar contains not less than four grams of acetic acid (CH_3COOH) per 100 milliliters The production of vinegar in the United States exceeds 100,000,000 gallons annually. It is used in the preparation of pickled fruits and vegetables or other large-scale food manufacture such as salad dressings, and as a seasoning agent for table use.

Types. Vinegars are classified as to raw materials used in their manufacture; there are products designated as wine, cider, malt, and sugar vinegar Wine vinegar is the product obtained from the acetous souring of grape wine. Cider vinegar is derived from the alcoholic fermentation of other fruit juices, usually apple juice, followed by acetous fermentation Malt vinegar is obtained by the alcoholic fermentation of starchy materials, such as cereal grains or potatoes saccharified, or converted to sugar, by the enzymes of malt, followed by acetous fermentation Distilled [illegible] usually belong to the malt vinegar c[illegible] egar is ob-



See RESONANCE (MOLECULAR STRUCTURE).

[R C FU.]

Viomycin

An antibiotic produced by an actinomycete to which the name *Streptomyces puniceus* has been applied (*see* ANTIBIOTIC; STREPTOMYCETACEAE). It is used in the sulfate form, which is prepared by conversion of an intermediate crystalline salt of the compound to its sulfate. Viomycin, like streptomycin, isoniazid (isonicotinic acid hydrazide), and *para*-aminosalicylic acid, is employed widely in the treatment of tuberculosis, particularly during certain stages of the disease (*see* PARA-AMINOSALICYLIC ACID, ISONICOTINIC ACID HYDRAZIDE, STREPTOMYCIN, TUBERCULOSIS)

Streptomyces puniceus was originally isolated from soil It derives its name from the purplish-red color of its vegetative mycelium. It grows readily under submerged aerobic conditions, and when cultivated in a medium containing suitable carbon and nitrogen sources and certain inorganic salts, it produces the antibiotic, viomycin Viomycin is excreted from the organism into the culture medium and may be recovered from the medium by filtration, absorption on ion exchange resin, elution with aqueous sulfuric acid and subsequent crystallization by the addition of methanol (*see* ION EXCHANGE) The product is a strong organic base and forms essentially neutral salts The sulfate, hydrochloride, and oxalate have been prepared in crystalline form Hydrated viomycin sulfate decomposes at 280°C. while its specific optical rotation is −32° (*see* OPTICAL ACTIVITY) It is highly soluble in water, and virtually insoluble in all organic solvents It is moderately stable in aqueous solution at pH 5–6, at room temperature Viomycin has a high molecular weight, and gives the following analysis carbon, 37 19%, hydrogen, 5 86%, nitrogen, 20 61%, and sulfate, 17 63% Acid and basic hydrolysates of the compound are biologically inactive

Viomycin inhibits the growth of many microbial species, including certain gram-negative bacteria and the mycobacteria Unlike other antibiotics it is relatively more active against the mycobacteria than against other species It is particularly active against *Mycobacterium tuberculosis*, the etiologic agent of tuberculosis Viomycin inhibits the growth of both strept[illegible] resistant [illegible] susceptible [illegible] though less active than streptomycin and isoniazid, it is more active than *para* aminosalicylic acid Drug-resistant strains of tubercle bacilli develop with viomycin, as with streptomycin and isoniazid, but cross resistance between these drugs has not been observed.

On administration to animals and man, viomycin is rapidly absorbed into the circulatory system and distributed throughout the body. It is excreted in high concentration in the urine. When administered in high dosage over prolonged periods of time, certain toxic manifestations of the drug have at times been observed. In particular, allergic reactions, abnormal renal function in individuals with preexisting renal damage, and occasional partial loss of hearing may occur.

The toxicity of viomycin is in general related to dosage. When used intermittently in low dosage, toxic reactions occur rarely.

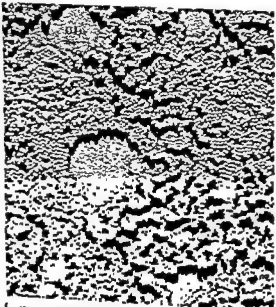
Viomycin is most effective therapeutically when used in combination with another antimicrobial drug It is used primarily in the treatment of patients whose infecting organisms are resistant to other antituberculous drugs and during resectional surgery of individuals with pulmonary tuberculosis Perhaps due in part to increased use of surgery in the treatment of this form of the disease, the use of viomycin has increased steadily since its introduction as a new drug in 1953 See ANTIMICROBIAL AGENTS, CHEMOTHERAPY

[G.L.H]

Bibliography. T. H Haskell, S A. Fusari, R. P. Frohardt, and Q. R Bartz, The chemistry of viomycin, *J. Am Chem Soc.*, 74,599–602, 1952; H Welch, *Principles and Practice of Antibiotic Therapy*, 1954

Virales

An order of minute, filterable agents of disease in higher animals, plants, and invertebrates from in-



[illegible] micrograph of purified polio virus (*E Ribi and B H Hoyer, Rocky Mountain Laboratory, USPHS*)

vinegar generator may operate satisfactorily for several years without loss of efficiency.

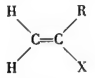
On the completion of fermentation of a batch of vinegar for table use, it is aged in well-filled, airtight wooden barrels or tanks for several weeks, then filtered, bottled, and pasteurized. Vinegars for pickling or other large-scale food manufacture may be filtered and pasteurized immediately after the fermentation is completed.

Continuous process. The use of an air-incorporation principle, known as a cavitator, in the generator has made possible continuous production of vinegar with 98% efficiency. The beechwood shavings are eliminated and the mash is aerated by a rotor and by air which is drawn down through the hollow shaft when rotor speed exceeds a predetermined value. The intimate air-mash mixture, produced by the rotor's nozzles, spreads out toward the tank's walls before rising. When it reaches the surface of the generator, the mash is recycled by going down through the draft tube to combine with air. Vinegar generation produces an increase in temperature which is controlled by cooling coils. *See* INDUSTRIAL MICROBIOLOGY. [L.B.LO.]

Bibliography: R. F. Cohee and G Steffen, *Food Engineering*, 31(3):58-59, 1959; S C. Prescott and C. G. Dunn, *Industrial Microbiology*, 2d ed., 1949, L. A Underkofler and R. J Hickey (eds), *Industrial Fermentations*, vol. 1, 1953.

Vinyl resin

A term that generally refers to addition polymers the monomers of which may be represented by the structure



where R and X may be hydrogen atoms or other substituents. *See* POLYVINYL RESINS.

[J A M , L M H]

Vinylogy

A principle used to correlate the properties of organic chemical compounds Organic compounds that differ from each other by a vinylene linkage (—CH=CH—) are said to be vinylogs of one another. Thus, ethyl crotonate is a vinylog of ethyl acetate and of the next higher vinylog, ethyl sorbate.

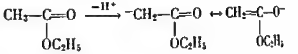
$$CH_3{-}CO_2C_2H_5$$
Ethyl acetate

$$CH_3{-}CH{=}CH{-}CO_2C_2H_5$$
Ethyl crotonate

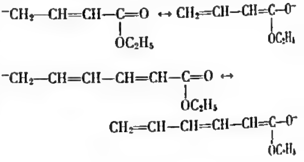
$$CH_3{-}CH{=}CH{-}CH{=}CH{-}CO_2C_2H_5$$
Ethyl sorbate

These esters are the first members of the vinylogous series $CH_3{-}(CH{=}CH)_n{-}CO_2C_2H_5$ The relationship between the members of such a series is known as vinylogy.

The outstanding characteristic common to the members of a vinylogous system is that the influence of the terminal functional groups upon each other persists throughout the series The readiness with which the methyl group of ethyl acetate loses a proton is ascribed to the enhanced stability of the resulting anion because of the presence of the ester group. Two principal resonance structures of the ion may be written



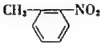
The anions (enolates) of ethyl crotonate and ethyl sorbate can be formulated in a similar way.

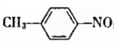


These enolates owe their stability in part to the resonance made possible by the vinylene linkages.

Similar relationships hold for aromatic systems in which two functions are situated in positions ortho or para to one another. Thus, the reactivities of the methyl groups in *o*- and *p*-nitrotoluene are similar to that of the methyl group in nitromethane, the first member of the vinylogous series

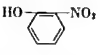
$CH_3{-}NO_2$ Nitromethane

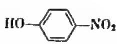

o-Nitrotoluene


p-Nitrotoluene

The compounds *o*- and *p*-nitrophenol, which may be described as vinylogs of nitric acid, are in fact much stronger acids than phenol itself.

$HO{-}NO_2$ Nitric acid


o-Nitrophenol

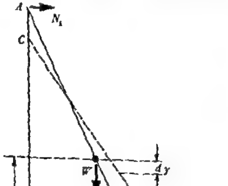

p-Nitrophenol

The principle of vinylogy may be used to explain why *o*- and *p*-nitrochlorobenzene are readily hydrolyzed to the corresponding nitrophenols, whereas the meta isomer is not.

That is, the principle of virtual work can be derived from these conditions, and conversely. *See* EQUILIBRIUM OF FORCES; STATICS

One advantage of the principle of virtual work is that it can serve as a basis for all of statics. In the solution of problems the principle of virtual work is often useful for eliminating the need for consideration of the forces of constraint, since these forces often are perpendicular to the virtual displacements and consequently do no work.

The following problem provides an application of the principle of virtual work (see illustration).



Illustrating the principle of virtual work (From R A Becker, Introduction to Theoretical Mechanics, McGraw-Hill, 1954)

A uniform plank of weight W rests on a smooth floor and leans against a smooth wall. The angle between the plank and the floor is ϕ. A weightless inextensible string connects the lower end of the plank to the wall. The value of the tension T in the string is desired. Let $2b$ be the length of the plank. Then as can be seen from the sketch,

$$y = b \sin \phi \qquad x = 2b \cos \phi$$

When ϕ increases by $d\phi$ the resulting changes dx and dy in x and y respectively are given by

$$dy = b \cos \phi \, d\phi \qquad dx = -2b \sin \phi \, d\phi$$

According to the principle of virtual work, the total work done in the virtual displacement produced by $d\phi$ must be zero. The forces of reaction N_1 and N_2 are normal to the wall, since the wall is smooth. Consequently they do no work. The total work, therefore, is the work done by the forces W and T moving through their respective virtual displacements dy and dx. Thus ...

$$0 = -2bT \sin \phi \, d\phi + Wb \cos \phi \, d\phi$$

Solving for the tension T gives

$$T = (W/2) \cot \phi$$

The same result is obtained from the equilibrium conditions expressed in terms of forces and torques. However, when the principle of virtual work is used, the forces of reaction need not be considered.
[P.W.S.]

Bibliography: R A Becker, *Introduction to Theoretical Mechanics*, 1954

Virulence

The capacity of microorganisms to inflict injury or death on a susceptible host. While one speaks of virulent microorganisms, actually this concept has no absolute, independent existence as a microbial attribute. The same microorganism may under identical conditions kill one host and be harmless to another; thus, the same microorganism is either virulent or avirulent. It follows that virulence acquires reality only when both the microorganism and the host are considered. Certain ancillary considerations contribute to the success or failure of the parasite. These are the numbers of the invading microorganism, the portal of entry, the mode of dissemination in the tissues and fluids of the host, the localization of the microorganism in certain tissues of the host, and the natural and induced humoral and cellular defensive reactions of the host. While virulence and pathogenicity are essentially synonymous, it has been proposed that the former term be applied to microbial species, types and strains, and the latter to broader taxonomic groupings, such as genera.

Microbial contributions to virulence. Given a susceptible host, microbial factors of virulence show many similarities but also significant differences among the major parasitic groups of viruses, rickettsia, bacteria, fungi, protozoa, and helminths. For example, in the case of the worms, even direct physical injury to the host acquires importance, as in the blockage of the bile duct or perforation of the wall of the intestine by the large roundworm, *Ascaris lumbricoides*. The bacteria however, to which the present discussion will largely be limited, injure the host not by sheer numbers nor by physical means, but primarily by their chemical activity. The bacteria may injure directly or may initiate a chain of secondary reactions. Principal microbial contributions to virulence are illustrated below, but it must be emphasized that more often than not one is concerned with the conjoint action of multiple factors rather than with a single injurious principle. *See* BACTERIOLOGY, MEDICAL; IMMUNITY; MYCOLOGY, MEDICAL; PARASITOLOGY, MEDICAL; RICKETTSIOSES; VIRUS

Exotoxins. These are most powerful poisons, protein in nature, readily released by the cells of a few bacteria, and they largely determine the course of such diseases as tetanus, botulism, gas gangrene, diphtheria, and staphylococcal food poisoning

sects down to bacteria The order has been assigned to the Microtatobiotes

The members of the order attacking the bacteria are called bacteriophages Included in the Virales are the microorganisms that cause such well-known human diseases as poliomyelitis, smallpox, and yellow fever, and the devastating foot-and-mouth disease and rinderpest of animals. Some of them are transmitted by insect and acarid vectors In general, susceptibility of viral agents to antibiotics has not been demonstrated (*see* POLIOMYELITIS; SMALLPOX; YELLOW FEVER).

Although systems of binary nomenclature in various degrees of inclusiveness of the above hosts have been proposed, none has been universally accepted as yet for classification of the Virales. In many groups of viruses, species are at present referred to by numbers, letters, or vernacular and even unauthored, latinized names. Agreement on classification of the viruses of insects is farthest advanced, but until general agreement is reached, it is not possible to review any formal system of classification in the Virales.

The demonstration by B. H. Hoyer in 1958 of the differential behavior of viruses, as well as rickettsiae, in cellulose ion exchange columns and by other chromatographic procedures should greatly assist in their classification. *See* MICROTATOBIOTES, VIRUS. [C.B.P]

Vireo

Any of 41 species of New World perching birds of the family Vireonidae, closely related to the wood warblers One genus, *Vireo*, occurs in the United States, represented by 12 species.

The vireos, sometimes called greenlets, are nearly all olive-green birds dwelling either in the tree tops or in thick underbrush. They are scanners, but are more deliberate in their movements than the warbler and have slightly heavier bills Best known is the widely distributed red-eyed vireo, *Vireo olivaceus*, which nests throughout North America, except on the Pacific Coast, from southern Canada into northern Mexico. Its deliberate, repeated song of short phrases, somewhat like that of the robin though softer and more disconnected, has earned it the common name of preacher bird. Many of the vireo songs are variants of this pattern. The red eyed vireo is a tree-top bird rather common in woodlands, orchards, and urban shade trees

All vireos build delicate, woven nests like those of the oriole, wedged in the forks of a tree or shrub *See* PASSERIFORMES. [J.B.R]

The white-eyed vireo, *Vireo griseus;* length to 5½ in It stays hidden more than the red-eyed vireo. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Virgo

In astronomy Virgo is a zodiacal constellation handed down from antiquity. It is visible throughout the summer months. This sixth sign of the Zodiac represents a maiden in a half reclining position. It is identified with the goddess of justice, Astraea. The balance is seen by her side in the sky (*see* LIBRA). The brightest star in the constellation, Spica, is a spectroscopic binary with a massive dark companion. It is a fine first-magnitude star of the purest white tint, and also a navigational star. *See* CONSTELLATION. [C.S.Y]

Virial equation

In thermodynamics, an empirical equation of state with additional terms beyond those for an ideal gas. The additional terms or virial coefficients account for some of the differences between real gases and ideal gases (*see* VAN DER WAALS EQUATION). For an ideal gas $pv = R_uT$, where p is pressure, v is volume, R_u is the universal gas constant 1544/M, where M is molecular weight, and T is thermodynamic or absolute temperature. To account for departures by real gases from this idealized relation, the equation of state can be written

$$pv = A + \frac{B}{v} + \frac{C}{v^2} +$$

where A, B, C, are functions of temperature; for a van der Waals gas they are

$$A = R_uT$$
$$B = R_uTb - a$$
$$C = R_uTb^2$$

where a and b are the van der Waals constants. Another virial equation, one that fits experimental data well over a wide range of conditions, is the Beattie-Bridgeman equation. *See* BEATTIE-BRIDGEMAN EQUATION. [C.A.H]

Virtual work, principle of

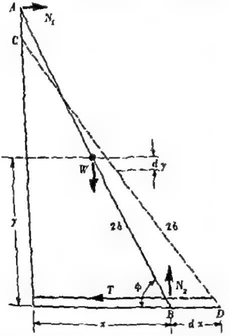
This principle states that the total virtual work done by all the forces acting on a system in static equilibrium is zero for a set of infinitesimal virtual displacements from equilibrium. The infinitesimal displacements are called virtual because they need not be obtained by a displacement that actually occurs in the system The virtual work is the work done by the virtual displacements, which can be arbitrary, provided they are consistent with the constraints of the system. *See* CONSTRAINT.

The principle of virtual work is equivalent to the conditions for static equilibrium of a rigid body expressed in terms of the tot s and torques

That is, the principle of virtual work can be derived from these conditions, and conversely. *See* EQUILIBRIUM OF FORCES; STATICS.

One advantage of the principle of virtual work is that it can serve as a basis for all of statics. In the solution of problems the principle of virtual work is often useful for eliminating the need for consideration of the forces of constraint, since these forces often are perpendicular to the virtual displacements and consequently do no work.

The following problem provides an application of the principle of virtual work (see illustration).



Illustrating the principle of virtual work (From R. A. Becker, Introduction to Theoretical Mechanics, McGraw-Hill, 1954)

A uniform plank of weight W rests on a smooth floor and leans against a smooth wall. The angle between the plank and the floor is ϕ. A weightless, inextensible string connects the lower end of the plank to the wall. The value of the tension T in the string is desired. Let $2b$ be the length of the plank. Then, as can be seen from the sketch,

$$y = b \sin \phi \qquad x = 2b \cos \phi$$

When ϕ increases by $d\phi$, the resulting changes dx and dy in x and y respectively are given by

$$dy = b \cos \phi \, d\phi \qquad dx = -2b \sin \phi \, d\phi$$

According to the principle of virtual work, the total work done in the virtual displacement produced by $d\phi$ must be zero. The forces of reaction N_1 and N_2 are normal to the wall, since the wall is smooth. Consequently they do *no* work. The total work, therefore, is the work done by the forces W and T moving through their respective virtual displacements dy and dx. This work is given by

$$0 = -2bT \sin \phi \, d\phi + Wb \cos \phi \, d\phi$$

Solving for the tension T gives

$$T = (W/2) \cot \phi$$

The same result is obtained from the equilibrium conditions expressed in terms of forces and torques. However, when the principle of virtual work is used, the forces of reaction need not be considered. [R.W.S.]

Bibliography: R. A. Becker, *Introduction to Theoretical Mechanics*, 1954.

Virulence

The capacity of microorganisms to inflict injury or death on a susceptible host. While one speaks of virulent microorganisms, actually this concept has no absolute, independent existence as a microbial attribute. The same microorganism may under identical conditions kill one host and be harmless to another; thus, the same microorganism is either virulent or avirulent. It follows that virulence acquires reality only when both the microorganism and the host are considered. Certain ancillary considerations contribute to the success or failure of the parasite. These are the numbers of the invading microorganism, the portal of entry, the mode of dissemination in the tissues and fluids of the host, the localization of the microorganism in certain tissues of the host, and the natural and induced humoral and cellular defensive reactions of the host. While virulence and pathogenicity are essentially synonymous, it has been proposed that the former term be applied to microbial species, types, and strains, and the latter to broader taxonomic groupings, such as genera.

Microbial contributions to virulence. Given a susceptible host, microbial factors of *virulence* show many similarities but also significant differences among the major parasitic groups of viruses, rickettsia, bacteria, fungi, protozoa, and helminths. For example, in the case of the worms, even direct physical injury to the host acquires importance, as in the blockage of the bile duct or perforation of the wall of the intestine by the large roundworm, *Ascaris lumbricoides*. The bacteria, however, to which the present discussion will largely be limited, injure the host not by sheer numbers nor by physical means, but primarily by their chemical activity. The bacteria may injure directly or may initiate a chain of secondary reactions. Principal microbial contributions to virulence are illustrated below, but it must be emphasized that more often than not one is concerned with the conjoint action of multiple factors rather than with a single injurious principle. *See* BACTERIOLOGY, MEDICAL; IMMUNITY; MYCOLOGY, MEDICAL; PARASITOLOGY, MEDICAL; RICKETTSIOSES; VIRUS.

Exotoxins. These are most powerful poisons, protein in nature, readily released by the cells of a few bacteria, and they largely determine the course of such diseases as tetanus, botulism, gas gangrene, diphtheria, and staphylococcal food poisoning

Their action may be highly specific, so that botulism and tetanus toxins act directly on neuronal elements, and the principal toxin of *Clostridium perfringens*, a cause of gas gangrene, attacks the lecithin widely distributed in host cells. Of special interest is the evidence that diphtheria toxin blocks the synthesis of some components of the host's cytochrome system. The discovery that bacterial viruses harbored by toxigenic strains of *Corynebacterium diphtheriae* can confer on atoxic strains the capacity to elaborate toxin has opened up a challenging new field of investigation. See BOTULISM; DIPHTHERIA; FOOD POISONING, BACTERIAL; GANGRENE, GAS; TOXIN, BACTERIAL.

Endotoxins. These substances are lipopolysaccharide in nature and are found within many bacterial cells. They are released by bacterial cells only after their disruption, and contribute to the injury inflicted by such unrelated organisms as those causing typhoid fever, bacillary dysentery, and meningococcus meningitis. See BACILLARY DYSENTERY; MENINGITIS; TYPHOID FEVER.

Extracellular enzymes and activators. Of the hundreds of microbial enzymes, only a few have to date been incriminated in microbial pathogenicity. One of these, hyaluronidase, or the spreading factor of F. Duran-Reynals, hydrolyzes the hyaluronic acid of the ground substance binding tissue cells together, and thus changes this tissue from a barrier to a pathway for microbial spread. An extracellular substance of pathogenic streptococci, streptokinase, activates the precursor of a proteolytic enzyme of plasma, which now can dissolve fibrin and remove it as a possible barrier to the dissemination of the streptococci. See STREPTOCOCCUS.

Surface components and phagocytosis. By significantly interfering with phagocytosis, substances such as the capsular polysaccharides of the pneumococcus make possible the proliferation of the microbes in the host. In a sense, they serve as tickets of admission to the host. See PNEUMOCOCCUS.

Microbial factors. These stimulate the release or formation of injurious substances by the host. There is substantial evidence that some of the most universal phenomena of infectious disease, such as fever and leukocytosis, are engendered by this sequence of events.

Contributions of host. The importance of the host in microbial virulence becomes apparent from the fact that many pathogens are harbored without overt disease, and conversely, many borderline opportunists can become deadly in a host whose resistance has been lowered. These opportunists have become increasingly troublesome, somewhat paradoxically, in the wake of some of the greatest modern medical advances, such as the antibiotics and cortisone. In addition to making possible the survival of hosts with lowered resistance, it is now recognized that these measures favor the opportunists by upsetting the balance of the indigenous microbial flora or by interfering with host response.

Host variables of significance in microbial pathogenicity are legion, and include such things as the genetic constitution, age, sex, nutrition, metabolism, hormonal state, other disease, fatigue, and specific hypersensitivity. As knowledge of the underlying mechanisms advances, many of these variables may be ultimately equated, in the sense that the effects of age or of sex may become explicable in terms of metabolic and hormonal factors. A beginning has been made toward determining the local biochemical factors at the site of host-parasite interaction to explain the fate of the parasite. It is known, for example, that the diabetic can fall victim to mucormycosis, caused by a generally benign mold. Experimentally it has been shown that host metabolic changes such as acidosis will significantly modify the invasiveness of the fungus. Similarly, in the tuberculous lesion the accumulation of keto-compounds favors the microorganisms, while locally produced lactic acid retards the microorganisms. It is anticipated that, with technological progress, further in vivo studies of the host-parasite interaction, both extra- and intracellularly, will lead to significant further progress in the understanding of the complex interplay of the two protagonists in microbial virulence. See IMMUNOLOGY. [M T]

Bibliography: H. Bauer, J. F. Flanagan, and W. H. Sheldon, The effects of metabolic alterations on experimental *Rhizopus oryzae* (mucormycosis) infection, *Yale J. Biol. Med.*, 29:23–32, 1956; R. J. Dubos, *Biochemical Determinants of Microbial Disease*, 1954; J. W. Howie and A. J. O'Hea (eds.), *Mechanisms of Microbial Pathogenicity*, 1955, W. D. McElroy and B. Glass (eds.), *The Chemical Basis of Heredity*, 1957; W. E. van Heyningen, *Bacterial Toxins*, 1950

Virus

An infectious agent mainly characterized by its small size and ability to reproduce only within living cells. Thus, a virus is too small to be seen with the ordinary microscopes used to visualize other microorganisms; and a virus will not reproduce in nutrient broth or jellies as other microbes do. A virus may multiply in appropriate tissue cultures because there it gets inside a cell as it does in the natural host.

Almost all forms of life are susceptible to virus infection including man, animals, plants, and even bacteria (bacteriophage). Viruses now cause enormous annual losses of domestic crops and animals, and they have been responsible for countless diseases since antiquity. In the case of man a virus was responsible for one of the three greatest scourges of disease in recorded history, the great influenza pandemic of 1918–1919, which is reported to have killed about 15,000,000 people. See ANIMAL VIRUS; BACTERIOPHAGE; PLANT VIRUS.

Size and shape. Direct visualization of a virus became common with the development of the electron microscope about 1940. Each virus was found to have a characteristic size and shape. Some are spheroidal bodies, some filament a others are loaf

shaped or polyhedral, and some are spermlike. All are exceedingly small, the largest being only a few hundred thousandths of an inch in diameter. Because of their small size and uniformity, particles of the simplest viruses are often referred to as molecules. Among the molecular properties shown by some viruses is crystallization. Tobacco mosaic virus was the first plant virus to be crystallized (W. M. Stanley, 1935) and polio virus was the first animal virus to be obtained in crystalline form (C. E. Schwerdt and F. L. Schaffer, 1955). Each virus crystal is composed of many thousands of virus particles. Virus preparations pure enough to crystallize usually provide the best material for chemical studies.

Chemical composition and structure. Each virus seems to have a definite and characteristic composition. The larger viruses tend to be more complex than the smaller ones both morphologically and chemically. Protein, nucleic acid, lipid, and carbohydrate have all been found in purified virus preparations. Few viruses contain all of these constituents, but all viruses appear to contain nucleic acid and protein. The simplest viruses consist solely of nucleic acid and protein in nucleoprotein combination. See CARBOHYDRATE; LIPID; NUCLEIC ACID; PROTEIN.

Electron micrographs of highly purified preparations of some viruses. (a) Vaccinia virus. (b) T2 bacteriophage. (c) T3 bacteriophage. (d) Tobacco mosaic virus. (e) Influenza virus. (f) Shope's rabbit papilloma virus. (g) Tomato bushy stunt virus. (h) Polio virus. Virus preparations by author and others, electron micrographs by R. C. Williams. (Virus Laboratory, University of California, Berkeley)

The protein components of viruses are somewhat acidic, have the usual amino acids found in all proteins, and are largely responsible for the morphology and serological properties of the viruses. A significant feature of viral protein structure is the presence of repeating subunits. Thus, the protein component of tobacco mosaic virus appears to consist of about 2100 identical peptide chains, each having a molecular weight of 18,000. These peptide subunits can be dissociated from the virus particles by treatment with such reagents as sodium dodecyl sulfate. While the proteins of different viruses have distinctive compositions, the proportions of amino acids in virus strains, also called variants or mutants, may or may not be different.

Viruses usually contain either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) located toward the interior of the virus particles and more or less closely associated with the peptide chains of the protein. The viruses of higher plants have RNA exclusively, while animal viruses may contain either RNA or DNA. Bacterial viruses seem to contain just DNA. The amounts of nucleic acid found range from somewhat less than 1% RNA in influenza virus to a little over 50% DNA in T2 bacteriophage. Different viruses usually have distinctive proportions of purine and pyrimidine components in their nucleic acids, whereas within strains of a virus, the proportions are the same. See DEOXYRIBONUCLEIC ACID; RIBONUCLEIC ACID.

The simpler viruses, at least, appear singularly lacking in common enzymatic components. Possible exceptions are influenza and avian myeloblastosis viruses for which neuraminidase and adenosine triphosphatase activities, respectively, have been reported.

Reproduction and mutation. The exact mechanism of virus duplication is obscure, but it appears that a single virus particle can initiate this process and that duplication is not by growth and division. Rather, new virus particles are synthesized in the infected cell from very small building blocks. There is evidence in some cases that this process can be initiated by something considerably less than an intact virus particle, perhaps by the nucleic acid alone.

Usually, reproduction results in the production of duplicate virus particles, but occasionally recognizable heritable changes appear in the offspring. This is variation or mutation and it is estimated to to occur with a frequency of 1 in 10^3 to 1 in 10^8 particles. Along with changes in the symptoms produced, mutation in certain plant viruses has been shown to be accompanied by demonstrable changes in the chemical structure of both the nucleic acid and the protein of the virus.

Relevance of virology to other areas. In addition to knowledge useful in the understanding and control of viral diseases, research on viruses is making a contribution in two other important areas, genetics and cancer.

The simpler viruses seem to present especially favorable material for genetic studies. These vi-

ruses can be isolated from host cells in a homogeneous condition, subjected to a variety of specific chemical and physical treatments, and then be returned to cells and the effects noted Such procedures are not readily applicable to most other genetic systems

Viruses have been shown to be responsible for the initiation of certain cancers in rabbits, frogs, fowl, mice, and plants. Until 1959 there has been little direct evidence to support a viral cause of human cancer. However, various characteristics of known cancer virus-host cell systems have suggested that a viral etiology of some human cancers is definitely possible, thus providing a fresh line of approach to this problem. *See* TUMOR VIRUSES.

Finally, the chemical investigation of the simpler viruses is providing a kind of information about protein and nucleic acid structures which should do much to clarify the role of these important substances in the vital processes of living systems. *See* AMINO ACIDS; MICROTATOBIOTES [C.A.K.]

Bibliography: F. M. Burnet and W. M Stanley (eds.), *The Viruses*, 2 vols., 1959; S. E Luria, *General Virology*, 1953.

Viscosity of gases

Viscosity denotes in general a resistance to flow Where velocity differences occur between groups of molecules, the mutual restraint caused by their interaction is called viscosity. This inner friction tends to reduce the velocity gradient by equalizing the two speeds, much as the jumping back and forth by passengers on two parallel trains moving at different speeds would tend to equalize their speeds. Viscosity is measured in terms of the applied force necessary to maintain a steady state of constant velocity difference between the two groups. Assuming two hypothetical planes of molecules having a common area A, a velocity difference dv, and a separation of r, the necessary applied force f is proportional to A times the velocity gradient dv/r, $f = \eta A dv/r$. The proportionality constant η is called the viscosity coefficient; in the cgs system it is given in units of poise, which have the dimensions of dyne-sec/cm^2 A coefficient of 1 poise will therefore require 1 dyne to sustain a relative flow rate of 1 cm/sec between molecule layers of unit cross section and 1 cm apart. *See* KINETIC THEORY OF MATTER

For gases one may assume an average distance between hypothetical layers equal to the mean free path l for that one-third of the n molecules/ml, of mass m, which may be presumed to move in a direction perpendicular to the layers The net effect of the molecules upon a layer is the average of their momentum transfer, that is, $\frac{1}{3}nml(dv/r)$. For a unit velocity gradient, this becomes the viscosity coefficient $\eta = \frac{1}{3}nmvl$. Because nm is the gas density, $\eta = \frac{1}{3}\rho vl$. The mean free path, l, may be determined from the molecular diameter σ and average speed to give

$$l = \text{speed/collision rate} = \frac{1}{\sqrt{2}\pi n\sigma^2}$$

Substituting for l in the viscosity expression, one obtains $\eta = \dfrac{mv}{3\sqrt{2}\pi\sigma^2}$ for gases.

This last expression for gas viscosity has two important implications. First is the independence of gas viscosity from its density. It might have been expected that the larger number of molecules present at high density would give more momentum transfer (viscosity), but the mean free path is so much decreased that momentum transfer is actually reduced. The second observation relates to temperature effect. Whereas a rise in temperature lowers the viscosity of a liquid, it increases the viscosity of a gas. This is shown by the dependence of η on v, which is proportional to $\sqrt{T}$. The difference in behavior is probably a result of the much stronger cohesive forces in a liquid, which are overcome at high temperature by increased kinetic energy. In gases a temperature rise results in more rapid transfer of momentum, hence in greater viscosity; in practice, the increase is somewhat greater than the $\sqrt{T}$ factor would indicate Gas viscosity, being much smaller than liquid viscosity, is of more theoretical interest than practical concern Its measurement is used to calculate the mean free paths of molecules. *See* FLUID-FLOW PROPERTIES; GAS; VISCOSITY OF LIQUIDS. [A.L.H.]

Viscosity of liquids

That property according to which liquids flow under the application of external stresses, but oppose the flow by internal forces. Viscosity, like diffusion and thermal conductivity, is one of the transport properties of liquids which have been extensively studied, both experimentally and theoretically, in efforts to develop theories of the liquid state. Its technical importance arises in lubrication and from the limitations it imposes upon the movement of fluids in pipes and in mixing operations.

Viscosity is defined as the force per unit area which resists the flow of two parallel fluid layers past one another when their differential velocity is 1 cm/(sec)(cm) separation The viscous force is described by Newton's equation

$$f = \eta A(dv/dx)$$

where f is the viscous force (dynes), A is the area (cm^2), (dv/dx) is the velocity gradient (sec^{-1}) and η is the coefficient of absolute viscosity (poise) The unit of absolute viscosity is the poise, equal to 1 dyne-sec/cm^2; but the smaller units, millipoise and centipoise, are more frequently used The coefficient of viscosity of water at 20°C is 10.087 millipoises, but the range of viscosities exhibited by liquids is very wide. Thus, the coefficient of viscosity of pentane is 2.395 millipoises at 20°C whereas that of glycerol is 10,690 millipoises at 20°C.

Liquids generally obey Newton's equation when he velocity gradient is not too great, and are said o be Newtonian in their flow characteristics In general, Newtonian fluids exhibit laminar or streamline flow. When the velocity exceeds a critical velocity in a tube of given radius, the flow becomes turbulent, and the viscosity is no longer given by Newton's equation The criterion for the onset of turbulent flow is that the quantity $2r\rho v/\eta$ exceeds the Reynolds number (2000), where r is the radius of the container, ρ is the density of the liquid, and v is the velocity of flow The quantity η/ρ, which is of importance in this expression, is called the kinematic viscosity and has the unit, stoke.

The laminar flow of fluids in long narrow tubes is described by Poiseuille's equation:

$$\eta = \frac{\pi P r^4 t}{8lV}$$

where r is the radius of the tube (cm), l is its length (cm), P is the pressure (dynes/cm²), and V is the volume (cm³) flowing in time t (sec). Poiseuille's equation is based on the assumption that the layer of liquid in contact with the tube wall is stationary It provides the basis of the absolute measurement of viscosity.

Measurement of viscosity. Relative measurements of viscosity can be made much more readily than absolute measurements by use of either the *Ostwald capillary viscometer* or the *falling sphere viscometer* The former, shown in Fig 2, consists of two glass bulbs separated by a length of capillary tubing Liquid is drawn up into the upper bulb, and the time required for its meniscus to fall between calibration marks above and below the upper bulb is carefully noted A similar measurement is made with a liquid of known viscosity. From Poiseuille's equation it follows that

$$\frac{\eta_1}{\eta_2} = \frac{\rho_1 t_1}{\rho_2 t_2}$$

where ρ_1 and ρ_2 are the densities of the two liquids and t_1 and t_2 are the times

The falling sphere viscometer is based upon Stokes' law for the frictional force on a spherical body of radius r falling with constant velocity in a fluid of viscosity η:

$$f = 6\pi\eta v r$$

This force is equal and opposite to the net force of gravity acting on the sphere,

$$f = \frac{4}{3}\pi r^3(\rho - \rho')g$$

where ρ and ρ' are the densities of a metal sphere and the fluid, respectively, and g is the acceleration of gravity Hence,

$$\eta = \frac{2gr^2(\rho - \rho')}{9v}$$

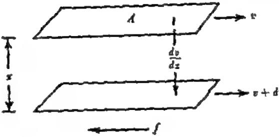


Fig 1. Viscous shear in liquids.

As with the Ostwald viscometer, it is simpler to compare the times of fall of the sphere in a fluid of unknown viscosity and in a fluid of known viscosity, and to use the relationship

$$\frac{\eta_1}{\eta_2} = \frac{t_1(\rho - \rho_1')}{t_2(\rho - \rho_2')}$$

Mechanism in viscosity. Although a number of empirical and theoretical relationships have been developed between viscosity and other properties of liquids, the fundamental mechanism of the transfer of momentum between molecules in a liquid undergoing viscous shear is still uncertain One empirical rule, known as Walden's rule, asserts that the product of the viscosity and the equivalent ionic conductance in electrolytic solutions should be a constant [illegible]

const [illegible]

rect. [illegible]

tions in the radii of the solvated ions in different solvents. *See* SOLUTION.

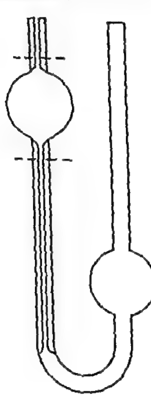


Fig 2 Ostwald viscometer.

Some nonassociated liquids conform well to an empirical relationship of the form

$$\eta = \frac{c}{v - \omega}$$

where v is the specific volume, and c and ω are parameters characteristic of the liquid. They may be evaluated from the measurement of η and v at two different temperatures. The value of ω is very closely equal to the van der Waals covolume b, and the term $v - \omega$ is thus approximately equal to the free volume of the liquid (*see* GAS). The proportionality of the fluidity (reciprocal of viscosity) and the free volume suggests that the free volume may be localized in the form of holes in the liquid structure. *See* MOLECULAR ASSOCIATION

A more nearly exact relationship describes the temperature-dependence of viscosity, $\eta = Ae^{B/RT}$, where A and B appear to be temperature-independent parameters characteristic of the liquid. The exponential form of this relationship has led to theories which describe viscous flow as a thermally activated process. It has been suggested that the configuration of molecules in the liquid state closely resembles that in the crystalline state for a very short distance around any given molecule. The term B represents the energy necessary to provide a hole of molecular dimensions in the liquid and to displace a neighboring molecule into it under the action of shearing stresses.

Hydrostatic pressure decreases the fluidity of liquids, in general. Significant experiments have been carried out on a number of liquids, in which viscosity has been measured at constant volume by varying pressure and temperature reciprocally. The results show that at constant volume, increasing temperature still decreases the viscosity, although the effect is much smaller than at constant pressure. The principal effect of temperature is to decrease the viscosity by increasing the volume of the liquid, but this is not the entire explanation. The hole theory of liquids is doubtless an oversimplification, and more precise measurements must be made before more accurate laws of liquid viscosity can be deduced. *See* FLUID MECHANICS, FLUID-FLOW PRINCIPLES; FLUID-FLOW PROPERTIES, LIQUID. [N.H.N.]

Bibliography: S. Glasstone, K. J. Laidler and H. Eyring, *The Theory of Rate Processes*, 1941; E. A. Moelwyn-Hughes, *Physical Chemistry*, 1957

Viscous flow

Viscosity represents the property of a fluid by which the fluid resists alteration of form. The effect of viscosity is predominant in determining such flow phenomena of a fluid as skin friction, form drag of a body, and lack of slip on the surface of a solid body. Viscous resistance is comparatively small for some fluids, such as water or gases, but large for other fluids, such as oil or glycerine. Some of the kinetic energy of flow is dissipated as heat through the viscous stresses.

For ordinary gases and liquids, called Newtonian fluids, viscous stresses may be considered to be proportional to the rate of change of velocity. The proportional factor is called the coefficient of viscosity μ. The coefficient of viscosity for liquids decreases as the temperature increases, but for gases, it increases with temperature and is independent of pressure when the pressure is moderate.

The most important parameter of a viscous flow is the Reynolds number $R_e = UL\rho/\mu$, where U and L are the characteristic velocity and length of the flow field, and ρ is the density of the fluid. For small Reynolds numbers, the viscous effect is important in the whole flow field, while for large Reynolds numbers the viscous effect will be important only in the boundary region. *See* BOUNDARY LAYER FLOW; REYNOLDS NUMBER. [S.I.P.]

Vision

The sense of sight, which perceives the form, color, size, movement, and distance of objects. Of all the senses, vision provides the most detailed and extensive information about the environment. Conversely, blindness is recognized as more disabling than deafness or any other sensory handicap. In the higher animals, especially the birds and primates, the eyes and the visual areas of the central nervous system have developed a size and complexity far beyond that of the other sensory systems.

Visual stimuli. These are typically rays of light entering the eyes and forming images on the retina at the back of the eyeball (Fig. 1). The intensity and wavelength characteristics of the light vary according to the light source and the object from which they are reflected (*see* COLOR; LIGHT). Human vision is most sensitive for light comprising the visible spectrum in the range from 380–720 mμ in wavelength. Sunlight and common sources of artificial light contain substantially all wavelengths in this range but each source has a characteristic spectral energy distribution. In general, light stimuli can be measured by physical means with respect to their energy, dominant wavelength, and spectral purity. These three physical aspects of the light are closely related to the perceived brightness, hue, and saturation respectively.

Atypical (sometimes called inadequate) stimuli for vision include momentary pressure on the eyeball, electric current through the eyes or head, a sudden blow on the back of the head, or disturbances of the central nervous system, caused by drugs, fatigue, or disease. Any of these may yield visual experiences not aroused by light. They are of interest because they show that the essentially visual character of the sensory experience is determined by the region stimulated (eyes, visual tracts) rather than by the nature of the stimulus. Indeed any observant person can detect swirling clouds or spots of "light" in total darkness or while looking at a homogeneous field such as a bright blue sky. These phenomena illustrate the spontaneous activity that is characteristic of the nervous system in general. that the vis-

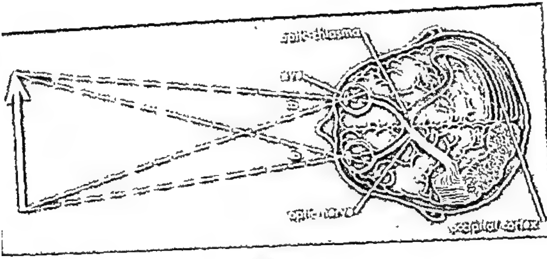


Fig. 1. Diagram showing the eyes and visual projection system

ual system is continuously active, which in turn means that the effect of a stimulus is to modify existing activity and not merely to initiate new activity.

Anatomical basis for vision. The anatomical structures include the eyes, optic nerves and tracts, optic thalamus and visual cortex. The eyes are motor organs as well as sensory; that is, each eye can turn directly toward an object to inspect it. The two eyes are coordinated in their inspection of objects, and they are able to converge for near objects and diverge for far ones. Each eye can also *regulate* the shape of its crystalline lens to focus the rays from the object and to form a sharp image on the retina. E...

... motor responses of the eyes are examples of involuntary action that is controlled by various reflex pathways within the brain. See EYE.

The process of seeing begins when light passes through the eye and is absorbed by the sensitive cells of the retina. These cells are activated by the light in such a way that electrical potentials are generated. These potentials are probably responsible for many features of the electroretinogram, an electrical response wave that can be detected by means of electrodes attached to the outside of the eye. These potentials are some of the signs of electrochemical activity that serves to trigger off nerve impulses in various successive neural cells (bipolars, ganglion cells and others) in the vicinity of excitation. Finally impulses emerge from the eye in the form of repetitive discharges in the fibers of the optic nerve. It must be emphasized, however, that the optic nerve impulses do not mirror exactly the excitation of the sensory cells by light. Complex interactions within the retina serve to enhance certain responses and to suppress others. Furthermore it is a fact that each eye contains more than a hundred times as many sensory cells as optic nerve fibers. Thus it would appear that much of the integrative action of the visual system has already occurred within the retina before the brain has had a chance to act.

The optic nerves from the two eyes meet at the optic chiasma. Figure 1 shows that the fibers from the inner (nasal) half of each retina cross over to the opposite side, while those from the outer (temporal) half do not cross over but remain on the same side. The effect of this arrangement is that the right visual field, which stimulates the left half of each retina, activates the left half of the thalamus and visual cortex. Conversely the left visual field affects the right half of the brain. This situation is therefore similar to that of other sensory and motor projection systems in which the left side of the body is represented by the right side of the brain and vice versa.

The visual cortex includes a projection area in the occipital lobe of each hemisphere (see BRAIN). Here there appears to be a point-for-point correspondence between the retina of each eye and the cortex. Thus the cortex contains a "map" or projection area, each point of which represents a point on the retina and therefore a point in visual space as seen by each eye. Nothing is known of the mechanism by which the two retinal maps are merged to form the cortical projection area. This merger allows the separate images from the two eyes to interact with each other in stereoscopic vision, binocular color mixture, and other phenomena. In addition to the projection areas on the right and left halves of the cortex there are visual association areas and other brain regions that are involved in vision. Complex visual acts, such as form recognition, movement perception, and reading are believed to depend on widespread cortical activity beyond that of the projection areas.

SCOTOPIC AND PHOTOPIC VISION

Night animals, such as the cat or the owl, have eyes that are specialized for seeing with a minimum of light. This type of vision is called scotopic. Day animals such as the horned toad, ground squirrel, or pigeon have predominantly photopic vision. They require much more light for seeing, but their day-

Some nonassociated liquids conform well to an empirical relationship of the form

$$\eta = \frac{c}{v - \omega}$$

where v is the specific volume, and c and ω are parameters characteristic of the liquid. They may be evaluated from the measurement of η and v at two different temperatures. The value of ω is very closely equal to the van der Waals covolume b, and the term $v - \omega$ is thus approximately equal to the free volume of the liquid (*see* GAS). The proportionality of the fluidity (reciprocal of viscosity) and the free volume suggests that the free volume may be localized in the form of holes in the liquid structure. *See* MOLECULAR ASSOCIATION.

A more nearly exact relationship describes the temperature-dependence of viscosity, $\eta = Ae^{B/RT}$, where A and B appear to be temperature-independent parameters characteristic of the liquid. The exponential form of this relationship has led to theories which describe viscous flow as a thermally activated process. It has been suggested that the configuration of molecules in the liquid state closely resembles that in the crystalline state for a very short distance around any given molecule. The term B represents the energy necessary to provide a hole of molecular dimensions in the liquid and to displace a neighboring molecule into it under the action of shearing stresses.

Hydrostatic pressure decreases the fluidity of liquids, in general. Significant experiments have been carried out on a number of liquids, in which viscosity has been measured at constant volume by varying pressure and temperature reciprocally. The results show that at constant volume, increasing temperature still decreases the viscosity, although the effect is much smaller than at constant pressure. The principal effect of temperature is to decrease the viscosity by increasing the volume of the liquid, but this is not the entire explanation. The hole theory of liquids is doubtless an oversimplification, and more precise measurements must be made before more accurate laws of liquid viscosity can be deduced. *See* FLUID MECHANICS; FLUID-FLOW PRINCIPLES; FLUID-FLOW PROPERTIES; LIQUID. [N H W.]

Bibliography: S. Glasstone, K. J. Laidler, and H. Eyring, *The Theory of Rate Processes*, 1941; E. A. Moelwyn-Hughes, *Physical Chemistry*, 1957

Viscous flow

Viscosity represents the property of a fluid by which the fluid resists alteration of form. The effect of viscosity is predominant in determining such flow phenomena of a fluid as skin friction, form drag of a body, and lack of slip on the surface of a solid body. Viscous resistance is comparatively small for some fluids, such as water or gases, but large for other fluids, such as oil or glycerine. Some of the kinetic energy of flow is dissipated as heat through the viscous stresses.

For ordinary gases and liquids, called Newtonian fluids, viscous stresses may be considered to be proportional to the rate of change of velocity. The proportional factor is called the coefficient of viscosity μ. The coefficient of viscosity for liquids decreases as the temperature increases, but for gases, it increases with temperature and is independent of pressure when the pressure is moderate.

The most important parameter of a viscous flow is the Reynolds number $R_e = UL\rho/\mu$, where U and L are the characteristic velocity and length of the flow field, and ρ is the density of the fluid. For small Reynolds numbers, the viscous effect is important in the whole flow field, while for large Reynolds numbers the viscous effect will be important only in the boundary region. *See* BOUNDARY LAYER FLOW; REYNOLDS NUMBER. [S.I.P.]

Vision

The sense of sight, which perceives the form, color, size, movement, and distance of objects. Of all the senses, vision provides the most detailed and extensive information about the environment. Conversely, blindness is recognized as more disabling than deafness or any other sensory handicap. In the higher animals, especially the birds and primates, the eyes and the visual areas of the central nervous system have developed a size and complexity far beyond that of the other sensory systems.

Visual stimuli. These are typically rays of light entering the eyes and forming images on the retina at the back of the eyeball (Fig. 1). The intensity and *wavelength characteristics* of the light vary according to the light source and the object from which they are reflected (*see* COLOR; LIGHT). Human vision is most sensitive for light comprising the visible spectrum in the range from 380–720 mμ in wavelength. Sunlight and common sources of artificial light contain substantially all wavelengths in this range but each source has a characteristic spectral energy distribution. In general, light stimuli can be measured by physical means with respect to their energy, dominant wavelength, and spectral purity. These three physical aspects of the light are closely related to the perceived brightness, hue, and saturation respectively.

Atypical (sometimes called inadequate) *stimuli* for vision include momentary pressure on the eyeball, electric current through the eyes or head, a sudden blow on the back of the head, or disturbances of the central nervous system, caused by drugs, fatigue, or disease. Any of these may yield visual experiences not aroused by light. They are of interest because they show that the essentially visual character of the sensory experience is determined by the region stimulated (eyes, visual tracts) rather than by the nature of the stimulus. Indeed any observant person can detect swirling clouds or spots of "light" in total darkness or while looking at a homogeneous field such as a bright blue sky. These phenomena illustrate the spontaneous activity that is characteristic of the nervous system in general that the vis-

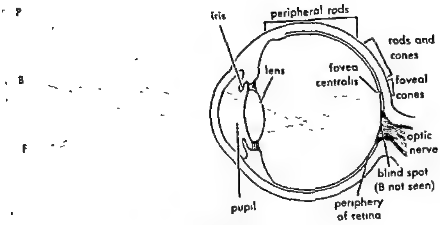


Fig 3 Foveal and peripheral vision Looking directly at F, the eye focuses the light from F on the central fovea, the region of clearest vision

look directly at a single letter at the center of a printed page This letter, and a few letters immediately adjacent to it, appear clear and black because they are seen with foveal vision The rest of the page is a blur in which the lines of print are seen as gray streaks. This is an example of peripheral vision Vision actually extends out to more than 90° from center, so that one can detect moving objects approaching from either side This extreme peripheral vision is like night vision in being devoid of sharpness and color

There is a simple anatomical explanation for the clarity of foveal vision as compared to peripheral vision The cone receptors become less and less numerous in the retinal zones that are more and more remote from the fovea In the extreme periphery there are scarcely any cones, and even the rod receptors are more sparsely distributed Furthermore the one-to-one neural connections from the foveal cones are replaced in the periphery by network connections in which hundreds of receptors may activate a single optic nerve fiber This mass action is favorable for the detection of large or dim stimuli in the periphery or at night, but it is unfavorable for visual acuity or color vision, both of which require that there be separate signals reaching the brain from the individual cone receptors.

Visual acuity. Visual acuity is defined as the ability to see fine details of an object In Fig 1 the small arrow at the back of each eye shows the image of the test object that is focused on the retina Standard visual acuity is defined as the ability to see an object so small that the angle Θ subtended at the eye is only 1 minute of arc, or 1/60 of a degree At 20 ft the size of such a test object is therefore only about 0 07 in., or 1 75 mm The image of the object on the fovea (neglecting diffraction and optical aberrations that deteriorate the image) has a length of only 0 005 mm or 5 μ. Small as this is, it is twice the diameter of the smallest foveal cone receptor One therefore comes to the conclusion that normal visual acuity approaches the limit imposed by the size and spacing of the receptors in the fovea Other limitations on visual acuity are imposed by diffraction and by the optical aberrations in the eye.

The specific forms of test object used for determining visual acuity yield different results. The angle Θ in Fig 1 can be made so small that it represents the size of test object that can barely be seen by the normal eye This angular size can be infinitely small in the case of a bright point seen against a dark background The stars at night provide a good example of this, since they are so distant that their angular subtense at the eye is practically zero A long dark line can be detected against a bright field (for example a flagpole seen against the sky) if it subtends one second of arc (1/60 of a minute) at the eye On the other hand, letters of the alphabet ...

eye A similar value holds for the lines of a grating (with parallel ...

single points or lines on the ...

nomenon resulting from the wave nature of light It means that no optical image can ever be completely sharp and clear The retinal image of a star is not a point but a blurred circle of light. The diameter of this blurred image is never less than 1 minute of arc, no matter how small an angle the star itself may subtend Thus the star is seen, provided that the blurred image is noticeably brighter

time vision is specialized for quick and accurate perception of fine details of color, form, texture, and location of objects. Color vision, when it is present, is also a property of the photopic system. Human vision is duplex; man is in the fortunate position of having both photopic and scotopic vision. Some of the chief characteristics of human scotopic and photopic vision are enumerated in the accompanying table.

Characteristic	Scotopic vision	Photopic vision
Photochemical substance	*Rhodopsin*	Cone pigments
Receptor cells	Rods	Cones
Speed of adaptation	Slow (30 min or more)	Rapid (8 min or less)
Color discrimination	No	Yes
Region of retina	*Periphery*	Center
Spatial summation	Much	Little
Visual acuity	Low	High
Number of receptors per eye	*120,000,000*	7,000,000
Cortical representation	Small	*Large*
Spectral sensitivity peak	505 mμ	555 mμ

Scotopic vision. This occurs when the rod receptors of the eye are stimulated by light The outer limbs of the rods contain a photosensitive substance known as visual purple or rhodopsin. This substance is bleached away by the action of strong light so that the scotopic system is virtually

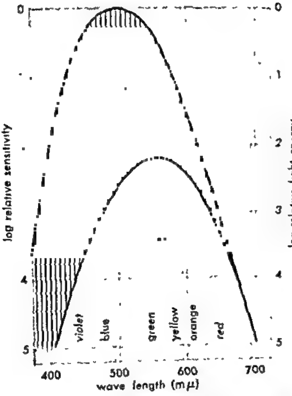


Fig. 2. Spectral sensitivity curves for human vision The rod curve shows that scotopic vision, based on rhodopsin, is most sensitive to light of about 505 mμ. The cone curve shows that photopic vision is generally less sensitive than scotopic vision, except for light at the red end of the spectrum.

blind in the daytime. In darkness, however, the rhodopsin is regenerated by restorative reactions based on the transport of vitamin A to the retina by the blood. One experiences a temporary blindness upon walking indoors on a bright day, especially into a dark room or dimly lighted theater As the eyes become accustomed to the dim light the scotopic system gradually begins to function. This process is known as dark adaptation Complete dark adaptation is a slow process during which the rhodopsin is restored in the rod receptors of the retina. Faulty dark adaptation or night blindness is found in persons who lack rod receptors or have a dietary deficiency in vitamin A. These rare persons are unable to find their way about at night without the aids of strong artificial illumination.

Dark adaptation is measured by an adaptometer, a device for presenting test flashes of light after various periods of time spent in the dark. The intensity of flash is varied to determine the momentary threshold for vision as dark adaptation proceeds A 10,000-fold increase in sensitivity (that is a reduction of 10,000 to 1 in the threshold intensity of flash) is often found to occur during a half-hour period of dark adaptation. By this time some of the rod receptors are so sensitive that only one elementary quantum (photon) of light is necessary to trigger each rod into action. A person can detect the presence of a flash of light that simultaneously affects only a few of the millions of rod receptors. Thus the scotopic sensitivity of the human eye is found to approach the ideal case of a receiver that is capable of responding to a single quantum of energy

The variation of the scotopic threshold with wavelength of light is shown in the rod curve of Fig 2. In spite of the variations in wavelength, the subject does not see any color when the intensity of the light is low enough to fall in the rod (shaded) portion of the diagram. This scotopic vision is colorless or achromatic, in agreement with the saying that in the night all cats appear gray

Normal photopic vision. Normal photopic vision has the characteristics enumerated in the table Emphasis is placed on the fovea centralis, a small region at the very center of the retina of each eye.

Foveal vision This is achieved by looking directly at objects in the daytime This means that the image of a small object falls within a region almost exclusively populated by cone receptors These are so closely packed together in the central fovea that their density is about 100,000 per square millimeter Furthermore, each of the cones in the fovea is provided with a "private line" of one-to-one nerve cells to the cortical projection area. The effect of this is that the cortex is supplied with superbly detailed information about any pattern of light that falls within the fovea centralis.

Peripheral vision. This is vision that takes place outside the fovea centralis (Fig 3). As an example,

of day, cloud cover, etc. The luminance range of objects seen on a given day is typically no more

the necessity for slowly adapting to the various levels of luminance.

Time perception. The temporal characteristics of vision are revealed by studying the responses of the eye to various temporal patterns of stimulation. When a light is first turned on, there is a vigorous burst of nerve impulses that travel from the eye to the brain. Continued illumination results in fewer and fewer impulses as the eye adapts itself to the given level of illumination. Turning the light off elicits another strong neural response. After-images are often seen at this time. A positive after-image, resembling the original stimulus, is sometimes seen during the first fraction of a second after the light goes off. This is usually followed by a longer lasting negative after-image in which the color of the original object appears to be reversed (see COLOR VISION). A clear after-image may be produced by staring fixedly at the stimulus for at least a half minute, then turning away and "projecting" the after-image against a white or gray screen. The retinal size of the after-image remains constant, so that its size in inches on a screen is proportional to the distance of the screen from the observer (Emmert's law). The after-image is thought to arise chiefly from the fact that the affected region of the retina has been bleached, or otherwise changed photochemically, in such a way that its responses are different from those of the regions not stimulated. To some extent, however, visual after effects are due to central rather than retinal, processes.

The strength of a visual stimulus depends upon its duration as well as its intensity. Below a certain critical duration the product of duration and intensity is found to be constant for a given amount of visual stimulation. A flash of light lasting only a few microseconds may stimulate the eye quite strongly ...
and ...
distu...
origi... luminance

Voluntary eye movements enable the eyes to roam over the surface of an object of inspection. In reading, for example, the eyes typically make four to seven fixational pauses along each line of print, with short jerky motions between pauses. An individual's vision typically takes place during the pauses, so that one's awareness of the whole object is the result of integrating these separate impressions over time.

A flickering light is one that is going on and off (or undergoing lesser changes in intensity) as a function of time. At a sufficiently high flash rate (called the critical frequency of fusion, cff) the eye fails to detect the flicker, and the light pulses seem to fuse to form a steady light that cannot be distinguished from a continuous light that has the same total energy per unit of time. As the flash rate is reduced below the cff, flicker becomes noticeable, and at very low rates the light may appear more conspicuous than flashes occurring at higher frequency. The cff is often used clinically to indicate a person's visual function as influenced by drugs, fatigue, or disease. See PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL. [L.A.R.]

Bibliography: Y. LeGrand, *Light, Colour and Vision*, 1957; M. H. L. Pirenne, *Vision and the Eye*, 1948.

Vitamin

An organic compound, essential for the normal functioning of the body, and usually obtained from foods. Vitamins are present in food in minute quantities compared to the other utilizable components of the diet—proteins, fats, carbohydrates, and minerals. The average adult eats about 600 grams of food per day on a dry weight basis, of which less than a gram are vitamins.

Almost all knowledge of the vitamins has been obtained during this century. The discovery of the vitamins has primarily been the result of two lines of investigation: the study of nutritional disease in people and the feeding of purified diets of known composition to experimental animals. In this way vitamin deficiency diseases, known as avitaminoses, have been described. The production of specific avitaminoses has usually been followed by the chemical synthesis of the protective vitamin after its isolation and characterization from natural food materials.

Synthetic and natural vitamins usually have the same biological value. The use of vitamins has been subject to many false claims by quacks and food faddists. The vitamins are not related chemically or functionally. They are conventionally divided into a ...
D, E, a...
of the v...
flavin, ...
folic ac... inositol, and
choline, as well as vitamin C, or ascorbic acid.

The vitamins, particularly the water-soluble ones, are found almost universally spread throughout the animal and plant kingdoms, functioning in essentially the same type of biochemical systems in the lowest and highest forms of life. Variations ...

... The vitamins function as coenzymes, active parts of enzyme systems, which catalyze many of the various anabolic and catabolic reactions of living organisms necessary for the production of energy, the synthesis of tissue components, hormones, and chemical regulators, and the detoxification and degradation of waste products and toxins. Because of their role in metabolism, the vitamins are generally concentrated

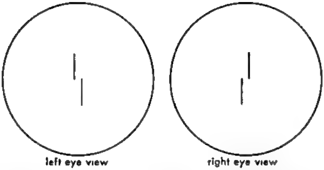


Fig. 4. Vernier and stereoscopic discriminations of space

than the surrounding field. In the case of a grating (black and white stripe pattern) the image of each line is blurred also. If the lines are too close (less than 1 minute) together the blurred image of one overlaps the blurred image of the next and the separate lines cannot be resolved by the eye. One comes to the conclusion, then, that the chief factors limiting the visual acuity of the fovea are (1) optical diffraction, (2) the ability to discriminate relative brightnesses within the images blurred by diffraction, and (3) the compactness of the pattern of cone receptors.

SPACE AND TIME PERCEPTION

Spatial and temporal effects are clearly apparent in the sense of sight. These enable the individual to orient himself with regard to space and time in the world around him, especially in the perception of motion and distance. *See* PERCEPTION.

Space perception. Elementary forms of space perception are vernier and stereoscopic discrimination. Here, the eye is required to judge the relative position of one object in relation to another (Fig. 4). The left eye, for example, sees the lower line as displaced slightly to the right of the upper. This is known as vernier discrimination. The eye is able to distinguish fantastically small displacements of this kind, a few seconds of arc under favorable conditions. If the right eye is presented with similar lines that are oppositely displaced, then the images for the two eyes appear fused into one and the subject sees the lower line as nearer than the upper. This is the principle of the stereoscope. Again it is true that displacements of a few seconds of arc are clearly seen, this time as changes in distance. The distance judgment is made not at the level of the retina but at the cortex where the spatial patterns from the separate eyes are fused together. The fineness of vernier and stereoscopic discrimination transcends that of the retinal mosaic and suggests that some averaging mechanism must be operating in space or time or both.

The spatial aspects of the visual field are also of interest. As has been previously indicated good acuity is restricted to a narrowly defined region populated by densely packed cones at the center of the visual field. A somewhat larger central region in which the cones are somewhat more sparsely distributed, is capable of good color vision. Farther out, however, vision is mediated by rod receptors; color vision is lost and form vision becomes extremely poor. In these peripheral regions area and intensity are reciprocally related for all small sizes of stimulus field. A stimulus patch of unit area, for example, looks the same as a patch of twice the same area and half the luminance. This high degree of areal summation is achieved by the convergence of hundreds of rod receptors upon each single optic nerve fiber. It is the basis for the ability of the dark-adapted eye to detect large objects even on a dark night.

In daytime vision spatial inhibition, rather than summation, is most noticeable. The phenomenon of simultaneous contrast is present at a border between fields of different color or luminance. Thus a small square of gray paper appears darker on a light background than on a dark one; on a yellow background it appears bluish. The effect may originate in neural mechanisms of inhibition such that stimulating a given region of a retina depresses the activity of regions immediately adjacent to it. This has the obvious effect of heightening contours and making forms more noticeable against their background (Fig. 5).

In photopic vision out of doors the eyes become light-adapted to an extent determined by the time

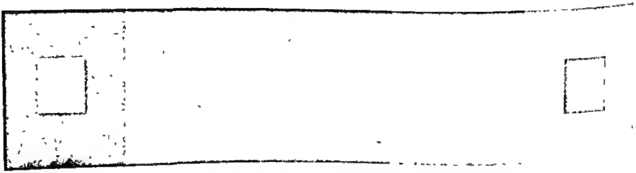


Fig. 5. Simultaneous contrast. The gray squares are physically equal and each strip is uniform. Spatial inhibitory effects cause the left square to look darker than the right square and cause the apparent darkening of the portion of each strip that is adjacent to a lighter strip. Thus the squares and the strip borders stand out more clearly.

common in many parts of the world and can be diagnosed by means of a biophotometer, which measures ability to adapt to changes in light. The role of vitamin A in sight has been related to a retinal pigment called rhodopsin, which is composed of a protein, opsin, and vitamin A. Under the stimulus of light, this pigment breaks down to opsin and vitamin A aldehyde, retinene. Resynthesis of rhodopsin, which is necessary for normal vision, is poor in vitamin A deficiency. The role of vitamin A in the metabolism of tissues other than the retina is unknown *See* VISION

Studies of many mammalian species have related the vitamin A requirement to body size Approximately 40 IU of β-carotene or 20 IU of vitamin A per kilogram of body weight will support growth and prevent symptoms of deficiency. In estimating human requirements, it has been assumed that two-thirds of the vitamin A consumed is in the form of carotenoid pigments There is evidence that it is of value to ingest considerably more vitamin A than the amount necessary to prevent deficiency signs. The recommended dietary allowances of the National Research Council are approximately twice the minimal requirements. These daily allowances vary from 1500 IU for 1- to 3-month infants to 5000 IU for normal adults and 8000 IU for lactating women Diseases which upset the normal digestion and absorption of the intestines or the presence of large amounts of mineral oil interfere with the absorption of the fat-soluble vitamins. *See* VITAMIN

[S.N.C.]

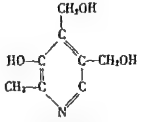
Vitamin A production. There are three approaches to the industrial preparation of vitamin A and its esters In one method, β-ionone is converted to the 14-carbon aldehyde (C_{14}-aldehyde), a key intermediate, by a glycidation reaction This aldehyde is condensed by an acetylenic Grignard reaction with a 6-carbon fragment, obtained by the ethynylation of methylvinyl ketone and by the rearrangement of the resulting carbinol. The condensed product is preferentially reduced, acetylated, dehydrated, and rearranged to give esters of vitamin A

The industrial synthesis of β carotene is based on the work of H H Inhoffen and F. Bohlmann and O. Isler and his coworkers In this method, the C_{14}-aldehyde, an intermediate in the vitamin A synthesis, is acetalized, then condensed with ethylvinyl ether and hydrolyzed to C_{16}-aldehyde This process is repeated using ethylpropenyl ether to give C_{19}-aldehyde Two moles of C_{19}-aldehyde are condensed with acetylene-dimagnesium bromide to yield a C_{40}-acetylenic glycol Dehydration, rearrangement, and preferential hydrogenation of this glycol yield β-carotene

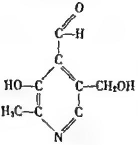
[A.O.]

Vitamin B_6

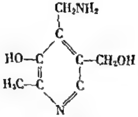
A vitamin which exists as three chemically related and water-soluble forms found in food, pyridoxine, pyridoxal, and pyridoxamine The structural formulas of these compounds are shown below. All three forms have equal activity for animals



Pyridoxine



Pyridoxal



Pyridoxamine

and yeast, but pyridoxal and pyridoxamine have several thousand times the activity of pyridoxine for some bacteria. Vitamin B_6 is stable in acid, but is rapidly destroyed by light in alkaline or neutral solutions

There are no satisfactory chemical assays for vitamin B_6 Although microbiological assays are possible, they often give results which appear much too high when compared to animal-feeding tests. This is because of the fact that vitamin B_6 is conjugated with other substances in foods, and the acid hydrolysis procedures used in preparing samples for microbiological assay release quantities of the vitamin not released by normal digestive processes. *See* BIOASSAY.

Most studies of vitamin B_6 deficiency have been done on animals, since a deficiency of this vitamin in humans is very rare Vitamin B_6 deficiency is accompanied by poor growth, dermatitis microcytic anemia, epileptiform convulsions, and kidney and adrenal lesions. There is evidence that some women in the third trimester of pregnancy may have a special requirement for vitamin B_6 in that its administration often relieves the nausea of pregnancy Some types of human dermatitis respond to local application of this vitamin

Vitamin B_6 functions as a coenzyme in the form of pyridoxal 5-phosphate. The value of pyridoxine and pyridoxamine lies in the ability of tissues to

Vitamin sources and deficiencies

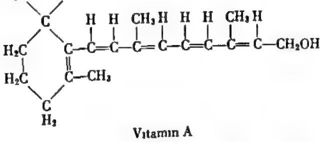
Vitamin*	Best sources	Deficiency	Areas primarily affected
Fat soluble			
α-, β- carotene	Fish-liver oils, green plants, carrots	Xerophthalmia, nyctalopia	Eyes, skin, mouth respiratory and urogenital tracts
D	Fish oils	Rickets, osteomalacia	Bones, teeth
E	Grain and vegetable oils		Reproductive tract, muscle, red blood cells, liver, brain
K	Green vegetables		Blood prothrombin
Water-soluble			
Thiamine	Pork, liver, whole grains	Beri beri	Brain, nerves, heart
Riboflavin	Milk, egg white, liver, leafy vegetables		Skin, mouth, eyes, liver, nerves
Niacin	Yeast, wheat germ, meats	Pellagra	Gastrointestinal tract, skin, brain
Pantothenic acid	Liver, kidneys, green vegetables, egg yolk		Adrenals, kidneys, skin, brain, spinal cord
B_6	Whole grains, yeast, egg yolk, liver		Skin, red blood cells, brain, kidneys, adrenals
Inositol	Whole grains, liver		
Biotin	Liver, kidney, yeast		Skin, muscle
Choline	Egg yolk, brain, grains		Liver, kidney, pancreas
p-Aminobenzoic acid	Yeast		
Folic acid	Liver, deep green leafy vegetables	Various macrocytic anemias	Red blood cells
B_{12}	Liver, meats	Pernicious anemia	Red blood cells
Ascorbic acid	Citrus fruits, fresh vegetables, potatoes	Scurvy	Bones, joints, mouth, capillaries

* See individual articles on each vitamin The lettered vitamins will be found as separate articles under vitamin

in those animal and plant tissues which are most active metabolically. Thus, liver or kidneys are a more potent source of vitamins than muscle or skin, and the germ of a seed contains more vitamins than its other parts *See* NUTRITION [S N C]

Vitamin A

A pale-yellow alcohol, soluble in fat, but not in water. It is readily destroyed by oxidation, which causes significant losses during storage and cooking. The structural formula of vitamin A is shown below.



Vitamin A

There are three sources of vitamin A activity. Vitamin A itself is found in all animal tissues. It is particularly concentrated in the liver and viscera, and the most important natural sources of the vitamin are the fish-liver oils. The livers of some fresh-water fish contain vitamin A_2, which differs slightly from vitamin A in structure. Plants contain a number of carotenoid pigments, which can be converted to vitamin A in the tissues of animals, such as in the rat intestinal wall. *See* CAROTENOID

The vitamin A activity of these carotenoids varies with their chemical structure. Beta-carotene has the most activity. One international unit (IU) of vitamin A has been set at 0 3 micrograms (μg) of vitamin A or 0 6 μg of β-carotene. This is confusing, since the efficiency of conversion of β-carotene to vitamin A becomes greater in a deficiency state. Although biological rat assays are used, vitamin A and carotene are usually determined chemically by spectrophotometric techniques.

In vitamin A deficiency, the epithelial tissues of many organs are affected. Growth failure occurs, particularly in the bones and teeth In young animals, this can result in neurological symptoms resulting from pressures on the central nervous system. Changes occur in the skin, mouth, respiratory tract, urogenital tract, and some glands. The changes in epithelial tissues increase the susceptibility of the deficient organism to infection. Vitamin A does not affect the bacteria. It protects the integrity of the mucous membrane Besides severe inflammation of the eyes, or xerophthalmia, vitamin A-deficient animals suffer from night blindness, a condition known as nyctalopia. This condition is

produce riboflavin, penicillin, or streptomycin, in 10,000- to 50,000-gal tanks in a medium of yeast extract, minerals, and sucrose or other carbohydrate for a period of 72–120 hours after inoculation. Optimum vitamin titers of 2–6 mg/liter are obtained under conditions of mild aeration, though the organism can grow anaerobically. *See* PROPIONIBACTERIACEAE.

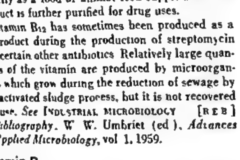
Since the vitamin is almost entirely contained in the cells, it can be recovered by centrifugation of the final broth. The cells can be dried, and used directly as a food or animal feed supplement. The product is further purified for drug uses.

Vitamin B_{12} has sometimes been produced as a by-product during the production of streptomycin and certain other antibiotics. Relatively large quantities of the vitamin are produced by microorganisms which grow during the reduction of sewage by the activated sludge process, but it is not recovered for use. *See* INDUSTRIAL MICROBIOLOGY [R E B]

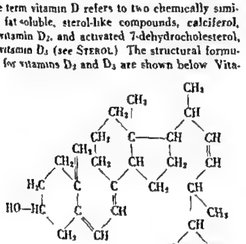
Bibliography. W. W. Umbriet (ed.), *Advances in Applied Microbiology*, vol. 1, 1959.

Vitamin D

The term vitamin D refers to two chemically similar fat-soluble, sterol-like compounds, *calciferol*, or vitamin D_2, and activated 7-dehydrocholesterol, or vitamin D_3 (*see* STEROL). The structural formulas for vitamins D_2 and D_3 are shown below. Vita-



Vitamin D_2 (calciferol)



Vitamin D_3 (activated 7-dehydrocholesterol)

mins D_2 and D_3 have about the same potency for mammals including man, but vitamin D_3 is 30–100 times as effective as vitamin D_2 for birds.

In animals, 7-dehydrocholesterol is secreted at the surf... ...re it is activated by sunlight, and the vitamin D_3 formed is reabsorbed. Vitamin D_2 is prepared by the ultraviolet irradiation of ergosterol, a vitamin D precursor found in plants. Most naturally occurring foods have little vitamin D activity. Fish oils are an exception, being highly potent sources of the vitamin. It is common practice to give children vitamin D supplements, and much of the milk marketed in the United States has been enriched with 400 IU per quart. An international unit (IU) of vitamin D is equal to 0.025 micrograms (μg) of vitamin D_3.

There are no satisfactory physical or chemical methods for the determination of vitamin D. Rats or chicks are used in biological assays. In these, measurements are made of the ash content of bones or beaks or the rate at which new bone is being formed. *See* BIOASSAY.

Vitamin D deficiency in growing animals is called rickets and has been produced in a large number of species. In adult animals, the disease is called osteomalacia. Vitamin D deficiency is associated with skeletal pathology. Bones and teeth are soft and fracture easily. Poor growth accompanied by malformations of the bones, particularly the long bones and the ribs, occurs. The disease is seldom fatal, but predisposes to other diseases, particularly bronchopneumonia.

Little is known of the metabolic role of vitamin D. It increases the intestinal absorption of calcium (Ca) but how it does this is not clear. It also appears to have an effect on bone formation not related to the absorption of Ca. In vitamin D deficiency, there is a failure of Ca salts to be deposited in the cartilaginous matrix of the bones.

The recommended daily dietary allowance of the National Research C...

...of the vitamin they obtain from their food to meet their needs. Because vitamin D is readily stored, single doses of 600,000 units two times a year, orally or intramuscularly, may be used if it is impractical to give children daily supplements or to enrich the milk. Massive doses of vitamin D have been used in the treatment of diseases having no relation to rickets, in some cases with favorable results. Doses of vitamin D 500–1000 times the recommended allowance, if continued over a long period of time, can result in extensive calcification of the kidneys, heart, arteries, stomach, and other soft tissues, accompanied by nausea, weakness, and even death. *See* VITAMIN [S N.C.]

Industrial production. There are three methods of enriching foods with vitamin D: (1) exposure to ultraviolet radiation, (2) feeding of irradiated yeast to dairy cattle, and (3) direct addition of a concentrate of vitamin D to the food. The third method is the only one extensively used.

Vitamin D is available in one of several forms. Irradiated yeast is the main source of vitamin D_2. Extraction of the yeast provides the provitamin ergosterol, which is subsequently irradiated and taken up in oil. Vitamin D_3 is obtained directly

transfer them into pyridoxal. Vitamin B_6-containing enzymes are involved in transamination reactions which are important mechanisms for the synthesis of amino acids by the tissues from α-keto acids. An example would be

glutamic acid + pyruvic acid ⇌
α-ketoglutaric acid + alanine

Vitamin B_6 is also involved in amino-acid decarboxylation reactions. The transformations of histidine to histamine and of aspartic acid to β-alanine are examples of these. This vitamin has a special function in the metabolism of tryptophan and is necessary for the conversion of tryptophan to niacin. In most vitamin B_6-deficient animals, xanthurenic acid, a metabolite of tryptophan, is excreted in abnormal quantities, and this has been used as the basis for tests of the adequacy of vitamin B_6 nutrition. This vitamin also has a special function in the metabolism of the sulfur-containing amino acids.

It is hard to set requirements for vitamin B_6, since no single set of assay conditions or criteria has received universal acceptance. Based on animal experiments, a number of dietary factors probably affect the vitamin B_6 requirement. High-protein diets increase the requirement. High-carbohydrate diets probably increase intestinal synthesis of vitamin B_6, while unsaturated fatty acids decrease the requirement. Adults probably require about 1–2 mg per day. There is evidence that infants 4–6 months old need 0.06–0.1 mg per day, or 0.01–0.02 mg/(kg)(day). *See* NIACIN; VITAMIN. [S.N.G.]

Pyridoxine production. Two prominent processes of the many industrial syntheses of vitamin B_6, called pyridoxine or adermin, are based on the publication of S. A. Harris and K. Folkers and the publications of other workers.

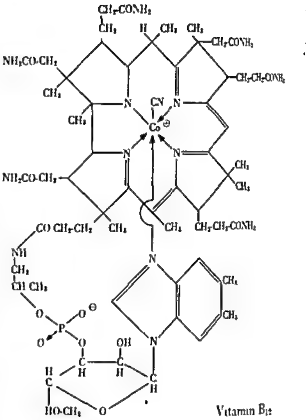
Both syntheses form pyridone rings by condensation of β-diketones with cyanoacetamide. With the first method, the ring is first nitrated in the 5 position, then the hydroxyl group replaced in 2 position by chlorine. In the second method, the ester in the 4 position is treated with ammonia to form the carbamide and dehydrated to the dinitrile, followed by nitration and chlorination. These compounds are then reduced to the corresponding di- or triamines which, upon diazotization, yield the vitamin. [M.C.]

Bibliography: S. A. Harris and K. Folkers, Synthesis of vitamin B_6, *J. Am. Chem. Soc.*, 61:1245–1247, 1939; J. H. Mowat, F. J. Pilgrim, and G. H. Carlson, *J. Am. Chem. Soc.*, 65:954–955, 1943.

Vitamin B_{12}

A group of closely related polypyrrole compounds containing trivalent cobalt. It is often called cobalamin, and its structural formula is given later. The vitamin is a dark-red crystalline compound, in aqueous solution and at room temperature it is most stable at pH 4–7.

In general, vitamin B_{12} is synthesized by microorganisms, not by plants, and is found in animal



Vitamin B_{12}

tissues as a result of intestinal synthesis or ingestion. It is now thought that the animal-protein factor, which was studied widely, is at least in part vitamin B_{12}. Vitamin B_{12} is usually determined microbiologically at the present time, although rat and chick growth assays are also available. *See* BIOASSAY.

Vitamin B_{12} deficiency in animals is characterized primarily by anemia. In man, this deficiency is called pernicious anemia. People suffering from this disease lack a factor secreted in normal gastric juice which, by affecting absorption directly or by protecting vitamin B_{12} from intestinal destruction, enables the vitamin to be absorbed. Vitamin B_{12} is concerned with methyl group metabolism and in this regard is probably often associated with folic acid. Requirements for vitamin B_{12} are increased by reproduction or hyperthyroidism. Of the known vitamins, B_{12} is the most active biologically. A daily injection of 1 μg of vitamin B_{12} will prevent the recurrence of symptoms in people with pernicious anemias. It would seem that the daily requirement of people must be less than 1 μg per day. In normal people, some if not all of this requirement is probably met by intestinal synthesis. *See* VITAMIN. [S.N.G.]

Vitamin B_{12} production. Processes for the commercial production of vitamin B_{12}, of which 1000 lb were produced in the United States in 1957 with a sales value of $21,869,000, utilize the synthetic ability of either bacteria or streptomycetes. Yeasts and higher fungi are, for the most part, nonsynthesizers, as are higher plants and animals.

Most commonly *Propionibacterium freudenreichii* or a similar organism is grown by the usual fermentation techniques, such as are employed to

produce riboflavin, penicillin, or streptomycin, in 10,000- to 50,000-gal tanks in a medium of yeast extract, minerals, and sucrose or other carbohydrate for a period of 72-120 hours after inoculation. Optimum vitamin titers of 2-6 mg/liter are obtained under conditions of mild aeration, though the organism can grow anaerobically. See PROPIONIBACTERIACEAE.

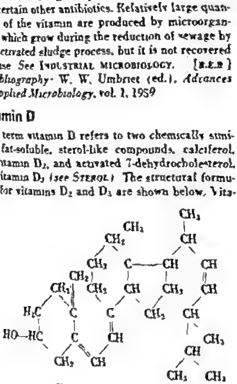
Since the vitamin is almost entirely contained in the cells, it can be recovered by centrifugation of the final broth. The cells can be dried, and used directly as a food or animal feed supplement. The product is further purified for drug uses.

Vitamin B_{12} has sometimes been produced as a by-product during the production of streptomycin and certain other antibiotics. Relatively large quantities of the vitamin are produced by microorganisms which grow during the reduction of sewage by the activated sludge process, but it is not recovered for use. See INDUSTRIAL MICROBIOLOGY. [R.E.B.]

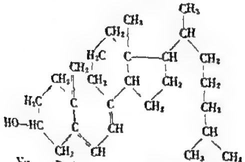
Bibliography: W. W. Umbreit (ed.), *Advances in Applied Microbiology*, vol. 1, 1959

Vitamin D

The term vitamin D refers to two chemically similar fat-soluble, sterol-like compounds, calciferol, or vitamin D_2, and activated 7-dehydrocholesterol, or vitamin D_3 (*see* STEROL). The structural formulas for vitamins D_2 and D_3 are shown below. Vitamins D_2 and D_3 have about the same potency for mammals including man, but vitamin D_3 is 30-100 times as effective as vitamin D_2 for birds.



Vitamin D_2 (calciferol)



Vitamin D_3 (activated 7-dehydrocholesterol)

In animals, 7-dehydrocholesterol is secreted at the surface of the skin, where it is activated by sunlight, and the vitamin D_3 formed is reabsorbed. Vitamin D_2 is prepared by the ultraviolet irradiation of ergosterol, a vitamin D precursor found in plants. Most naturally occurring foods have little vitamin D activity. Fish oils are an exception, being highly potent sources of the vitamin. It is common practice to give children vitamin D supplements, and much of the milk marketed in the United States has been enriched with 400 IU per quart. An international unit (IU) of vitamin D is equal to 0.025 micrograms (μg) of vitamin D_3.

There are no satisfactory physical or chemical methods for the determination of vitamin D. Rats or chicks are used in biological assays. In these, measurements are made of the ash content of bones or beaks or the rate at which new bone is being formed. See BIOASSAY.

Vitamin D deficiency in growing animals is called rickets and has been produced in a large number of species. In adult animals, the disease is called osteomalacia. Vitamin D deficiency is associated with skeletal pathology. Bones and teeth are soft and fracture easily. Poor growth accompanied by malformations of the bones, particularly the long bones and the ribs, occurs. The disease is seldom fatal, but predisposes to other diseases, particularly bronchopneumonia.

Little is known of the metabolic role of vitamin D. It increases the intestinal absorption of calcium (Ca) but how it does this is not clear. It also appears to have an effect on bone formation not related to the absorption of Ca. In vitamin D deficiency there is a failure of Ca salts to be deposited in the cartilaginous matrix of the bones.

The recommended daily dietary allowance of the National Research Council for children, pregnant, and lactating women is 400 IU. Normal adults can probably depend on sunshine and the small amount of the vitamin they obtain from their food to meet their needs. Because vitamin D is readily stored, single doses of 600,000 units two times a year, orally or intramuscularly, may be used if it is impractical to give children daily supplements or to enrich the milk. Massive doses of vitamin D have been used in the treatment of diseases having no relation to rickets, in some cases with favorable results. Doses of vitamin D 500-1000 times the recommended allowance, if continued over a long period of time, can result in extensive calcification of the kidneys, heart, arteries, stomach, and other soft tissues, accompanied by nausea, weakness, and even death. See VITAMIN. [S.N.G.]

Industrial production. There are three methods of enriching foods with vitamin D: (1) exposure to ultraviolet radiation; (2) feeding of irradiated yeast to dairy cattle, and (3) direct addition of a concentrate of vitamin D to the food. The third method is the only one extensively used.

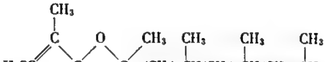
Vitamin D is available in one of several forms. Irradiated yeast is the main source of vitamin D_2. Extraction of the yeast provides the provitamin ergosterol, which is subsequently irradiated and taken up in oil. Vitamin D_3 is obtained directly

from marine sources, such as the fish-liver oils, or produced from animal tissues by extraction of the provitamin 7-dehydrocholesterol which is irradiated, and then taken up in oil.

Vitamin concentrates are prepared by dissolving the vitamins in milk carriers, which are canned and sterilized and used primarily for milk. Other carriers for vitamin D consist of propylene glycol, alcohol, oils with dispersing agents, or mixtures in compressed starch wafers. These are added directly to the foods to be fortified. Vitamin D_3 from fish-liver oils is the only naturally occurring form of vitamin D, and it is extracted by molecular distillation. *See* DISTILLATION, MOLECULAR. [K.G.W.]

Vitamin E

A group of compounds called tocopherols There are four of these, alpha (α), beta (β), gamma (γ), and delta (δ), each differing slightly in its chemical configuration. The structural formula of α-tocopherol is shown below. These structural differences have a decided effect on the biological activity of the compounds When the rat is used as a test animal, and the activity of α-tocopherol is set at 100, the activities of the β-, γ-, and δ-tocopherols would be 40, 8, and 1, respectively. In addition to the tocopherols, many synthetic compounds with similar structures and a number of antioxidants unrelated chemically to the tocopherols also have vitamin E activity



α-Tocopherol

The tocopherols are fat soluble They are widespread in nature and are particularly concentrated in vegetable oils. They are stable to heat, acids, and alkalies, but they are easily destroyed by ultraviolet light or oxidizing agents

Chemical and physical analyses for tocopherols are difficult, often expensive, and do not differentiate between the different tocopherols. Vitamin E activity can be determined biologically, but these methods are also disappointing *See* BIOASSAY.

Vitamin E deficiency in animals has often been associated with reproductive failure and irreversible testicular damage. This aspect of vitamin E deficiency has resulted in many unfounded claims as to the role of vitamin E in reproduction There is no good evidence that vitamin E deficiency is associated with any human reproduction problems. In vitamin E deficiency, dystrophy of both striated and cardiac muscles occurs, and death often occurs as a result of cardiac failure. Vitamin E therapy has not been of value in treating muscular dystrophy in humans. Vitamin E-deficient rats develop liver necrosis, and chicks develop encephalomalacia and subcutaneous edema The rat-liver necrosis and the edema of chicks can be cured by traces of selenium as well as vitamin E. The relation between vitamin E and selenium is not clear. The fat of some vitamin E-deficient animals becomes brown because of the deposition of ceroid pigments. The red blood cells of deficient animals are more subject to hemolysis, and this characteristic of vitamin E deficiency is used in biological assays for the vitamin. Practically nothing is known of how vitamin E functions in metabolism, although it seems reasonably certain that it functions as a biological antioxidant. In this capacity, it has been shown to have a sparing effect on the vitamin A content of tissues. There are no good data concerning the vitamin E requirements of humans The diet of the average adult contains approximately 20 mg of α-tocopherol per day *See* BIOLOGICAL OXIDATION; VITAMIN [S.N.G.]

Alpha-tocopherol production. Racemic *dl* α-tocopherol is synthesized industrially by development of a basic procedure described by P. Karrer in 1938. By this method, trimethylhydroquinone is condensed with phytyl bromide in the presence of zinc chloride The technical variations include the use of isophytol, phytol, and phytadiene with other catalysts. *See* CATALYSIS

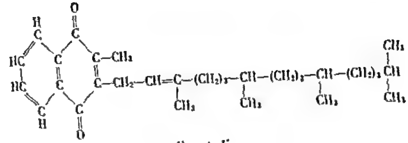
The product is purified by a short-path, high-vacuum distillation. The synthetic tocopherol is used mostly as the acetate, which is as active as the free phenol, but has greater stability toward oxidation Tocopherol acetate is manufactured by direct acetylation of free tocopherol with acetic anhydride, followed by a high-vacuum distillation *See* DISTILLATION, MOLECULAR. [A.O.]

Vitamin K

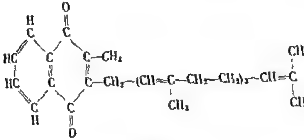
Naturally occurring vitamin K consists of two chemically similar yellowish oils. These vitamins, known as K_1 and K_2, are fat-soluble, nonsteroid, nonsaponifiable, and unstable to light. Many compounds with vitamin K activity have been synthesized The most important of these is menadione or K_3. The structural formulas for vitamins K_1, K_2, and K_3 are shown in the illustration.

Vitamins K_1, K_2, and K_3 have equal activities for most species when fed on an equimolecular basis, but not for the dog The only major effect of a vitamin K deficiency in animals is a decrease in blood prothrombin, resulting in a prolonged clotting time Thus, vitamin K deficiency results in subcutaneous and intramuscular hemorrhages The role of vitamin K in the production of prothrombin by the liver is not known It may act as a prosthetic group or activator for some enzyme necessary for the production of the prothrombin, or it may even be incorporated to some extent in the prothrombin molecule.

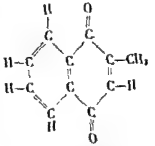
Vitamin K is so widespread in nature in green leafy vegetables and its s [illegible] by intestinal



Vitamin K_1



Vitamin K_2



Vitamin K_3 (menadione)

microorganisms is so great that vitamin K deficiency is rare. Liver disease, which may result in decreased *prothrombin* formation and bile secretion, can result in vitamin K deficiency, because bile secretion is necessary for its absorption. Intestinal disease or the ingestion of large amounts of mineral oil or antibiotics can also affect vitamin K economy. Relatively little is known about the biochemistry of vitamin K. It has been shown that vitamin K_1 is involved as a coenzyme in both electron transport and oxidative phosphorylation systems. *See* COENZYME.

There are no good chemical or physical assays for vitamin K. The vitamin is usually *determined* b...

... low prothrombin values are *often* given vitamin K therapy. The newborn infant has very low prothrombin values because of poor placental transfer, an undeveloped intestinal flora, and the low vitamin K content of milk. It is customary to give prospective mothers large doses of ...
or two befo...
fant some ...

[S.N.G.]

Vitamin K_1 production. The industrial synthesis of vitamin K_1 depends *upon the* condensation of phytyl moiety with 2-methyl-1,4-naphthoquinone. The phytyl moiety may be phytol or a phytyl halide.

In a typical synthesis, 2-methyl-1,4-naphthoquinone is converted ...
2-methyl-1,...
monobenzo...
is condense... phytyl moiety in the presence of an acid catalyst, such as zinc chloride, ...
oxidized to vitamin K_1. [A.O.]

Vivianite

The important minerals of the vivianite group are vivianite, annabergite, and erythrite.

Vivianite is a hydrated ferrous phosphate, $Fe_3(PO_4)_2 \cdot 8H_2O$. Usually ferric iron is present as the result of oxidation although this does not change the crystal structure appreciably. It crystallizes in the monoclinic system, with crystals generally prismatic. Vivianite also occurs in earthy form and as globular and encrusting masses of fibrous structure. Crystals are colorless and transparent when fresh. Oxidation changes the color progressively to pale blue, greenish blue, dark blue or bluish black.

Vivianite is widespread as a secondary *mineral* in gossans of metallic ore deposits, in *weathered* Mn-Fe phosphatic pegmatites, and in alluvial and sedimentary deposits associated with bone and other organic remains. As odontolite in fossil bone or teeth it is often ...

of ...
an ... to form a complete series between the end members $Co_3(AsO_4)_2 \cdot 8H_2O$ and $Ni_3(AsO_4)_2 \cdot 8H_2O$. Erythrite includes the half of the series with Co > Ni, while annabergite includes the half with Ni > Co. Both minerals crystallize in the monocl...
ra...
an...
oc...
ea... annabergite develops as finely

crystalline crusts and as earthy masses. Erythrite is crimson red and peach red. Increasing the Ni content changes the color progressively to pale pink, pale green, and apple green in annabergite

These secondary minerals are developed during oxidation of cobalt and nickel arsenides. Erythrite is the more common; both are found in Europe and North America. *See* COBALT. [W.R.LO.]

Vocal cords

The pair of ligamentous bands inside the human larynx. They act as sphincters for air regulation and may be vibrated to produce sound. The cords are covered with a mucous membrane. They pass horizontally backward from the thyroid cartilage (Adam's apple) to insert on the smaller, paired arytenoid cartilages at the back of the larynx. Separation, approximation, and alteration of tension are produced by action of laryngeal muscles acting on the pivoting arytenoids. Innervation is through branches of the vagus nerve. Vibration by air produces fundamental sounds and overtones. These can be modified by the strength of the air current, the size and shape of the glottis (the opening between the cords), and tension on the cords.

Among mammals, only man produces articulate speech, although the basic structures for sound production are similar in all In certain birds (for example, parrots) special modification of the larynx permits a degree of speechlike vocalization. *See* SPEECH. [E.G.ST.]

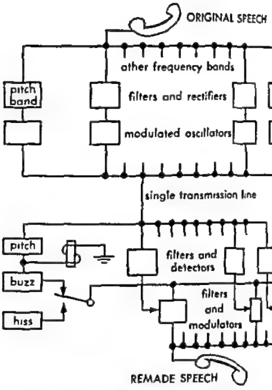
Vocoder

A system which analyzes speech for its essential elements, transmits these elements in a narrow frequency band to a distant point, and recreates the speech at the receiving point. The system was invented by H. W. Dudley at Bell Telephone Laboratories, Inc. The name is a contraction of the words voice coder.

To transmit speech intelligibly, a telephone or radio circuit must be capable of passing frequencies up to about 3000 cycles per second (cps). If a conversation can be compressed into a band only 300 cycles wide, for example, 10 such conversations can be transmitted simultaneously over one 3000-cycle channel, at considerable saving in cost The vocoder does this compressing by taking advantage of the fact that, although speech contains a wide band of frequencies, much of the information is contained in modulation of these frequencies at rates which may not exceed 25 cps.

At the transmitting end, speech is analyzed by filters into bands (10 or more have been used). The filter outputs are reduced to relatively slow fluctuations and are applied to modulate fixed tones at the desired transmitting frequencies. Each such modulated tone requires only about 25 cycles bandwidth. Another narrow band is used to transmit the slow fluctuations of pitch. All bands are sent over the same channel and sorted out at the receiving end

Two artificial sound sources are used for speech



Block diagram of a vocoder. (*After H Dudley*)

re-creation, a buzz (representing vocal cord tone) and a hiss (representing turbulent air noise) Both the switching between them and the frequency change of the buzz are controlled by signals in the pitch channel. One source or the other goes to another set of filters, whose output strengths are controlled by signals in the band channels. The resulting speech reproduction is intelligible. There is, however, an undesirable loss in naturalness which has prevented wide adoption of the system in telephony.

Subsequent to the invention of the system described, some other methods have been proposed which would seem to offer even greater reduction of bandwidth. One method utilizes the frequency changes within certain frequency regions (perhaps three) in the original speech spectrum and transmits this information for use in controlling speech re-creation Another attempts to recognize each speech sound of the original speech, code this information for transmission, and cause the same sounds to be produced artificially at the receiving point *See* SPEECH; *see also* FREQUENCY MODULATION. [H.K.D.]

Bibliography: H Dudley, Remaking speech, *J Acous. Soc. Am.* 11(2), 169–177, 1939.

Volatilization

The process of converting a chemical substance from a liquid or solid state to a gaseous or vapor state Other terms used to describe the same process are vaporization, distillation, and sublimation A substance can often be separated from another substance by volatilization and then recovered by condensation of the vapor. The substance can be made to volatilize more rapidly either by heating

increase its vapor pressure or by removal of the por using a stream of inert gas or a vacuum imp Heating procedures include the volatiliza-on of water, of mercury, or of arsenic trichloride separate these substances from interfering ele-ents Chemical reactions are sometimes utilized produce volatile products as in the release of rbon dioxide from carbonates, of ammonia in the jeldahl method for the determination of nitrogen, nd of sulfur dioxide in the determination of sulfur n steel Volatilization methods are generally char-cterized by great simplicity and ease of operation, xcept when high temperatures or highly corrosion-esistant materials are needed See DISTILLATION; SEPARATION (CHEMICAL AND PHYSICAL); SUBLIMA-TION; VAPOR PRESSURE. [L.G.]

Volcanic glass

A natural glass formed by rapid cooling of lava. The material is opaque except on thin edges and occurs in a variety of colors. Shades of red, brown, black, gray, or green may be displayed in a uniform banded, or variegated fashion Because of its conchoidal fracture (Fig 1), volcanic glass was highly prized among many primitive peoples as a material for tools and weapons

Fig 1. Specimen of volcanic glass showing smooth curved fracture surfaces High-silica lavas commonly solidify without crystallization because of the high viscosity and rapid chilling (Ward's Natural Science Establishment, Inc)

Types of natural glass. Most natural glasses are chemically equivalent to rhyolite into which they may grade Types corresponding to trachyte, dacite, andesite, and latite are much less common Basaltic glass generally is given the name tachylite The index of refraction and specific gravity of natural glass (see table) may aid greatly in its identification

Properties of natural glass

Type of glass	Average index of refraction	Average specific gravity
Rhyolitic	1 495	2.37
Trachytic	1 505	2 45
Dacitic and andesitic	1 515	2 50
Basaltic (tachylite)	1 575	2 77

Obsidian is generally a black variety with brilliant luster Pitchstone however has a dull or pitchy luster and is frequently brown, green, or gray Pumice is an amazingly light rock and consists of frothlike glass in which the tiny gas bubbles commonly may not exceed 1 mm in diameter Perlite is a gray to green glass with abundant spherical cracks (Fig 2a) which cause it to disintegrate into tiny pelletlike masses See OBSIDIAN, PERLITE, PITCHSTONE, PUMICE

Texture and structure. In addition to crystallites, most volcanic glasses carry microlites (microscopically tiny crystals) Many of these may show skeletal form Somewhat larger and usually well-formed crystals are called phenocrysts. Their presence gives the glass a porphyritic texture They consist chiefly of quartz, alkali feldspar, and plagioclase, and less commonly of mafics (biotite hornblende, or pyroxene). As the number of phenocrysts increases, these porphyritic glasses pass into vitrophyre (Fig. 2b). See PHENOCRYST; PORPHYRY.

Spherulites and lithophysae are developed strikingly in many glassy rocks. A common feature is fluidal structure resulting from flowage of the viscous lava It consists of wavy streaks and bands of spherulites, phenocrysts, microlites, and crystallites. In some glasses conspicuous banding is due to alternating layers of different colored glass or of glass and pumice.

Water content. The water content of natural glass is highly variable It is usually less than 1% in obsidian, between 3 and 4% in perlite, and between 4 and 10% in pitchstone The water in obsidian may

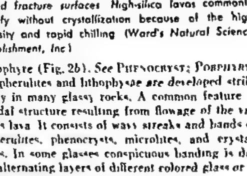


Fig 2 (a) Perlitic cracks in volcanic glass (b) Vitrophyre with phenocrysts of quartz and sanidine more or less oriented parallel to contorted flowage bands (fluidal structure)

represent only part of that contained in the original melt. Much water may have been lost as vapor when the lava was erupted. Some of the water in perlite and much of that in pitchstone is believed to have been derived by absorption from the sea or wet sediments into which the viscous lava came. Since glasses commonly are associated intimately with crystalline material, some of the water driven out of the crystallized portions may have been absorbed by the glassy portions.

Formation and composition. The formation of glass from magma (rock melt) is favored by rapid cooling and high viscosity. Rapid cooling is most readily attained at or near the earth's surface; consequently, glassy rocks are found as lava flows and shallow intrusives. High viscosity is characteristic of lavas rich in silica and potassium (rhyolitic); therefore, most glassy rocks are rhyolitic. Chemical composition may control the formation of glass in another respect. The composition of most natural glass is close to the *quartz–alkali feldspar cotectic.* This permits such lavas to cool to relatively low temperatures at which viscosity is high without crystallizing. If solidification can be forced upon the melt at this stage, crystallization will be impeded and glass will form. Such ideal conditions may be brought about by sudden eruption to the surface where chilling and loss of volatiles (fluxes) will occur. *See* FELDSPAR; LAVA; MAGMA

Glass tends to convert (devitrify) spontaneously to a crystalline aggregate. Devitrification commonly starts at the surfaces of phenocrysts and spherulites or along cracks, and it spreads outward. In time the entire mass of glass may be transformed to an aggregate of microscopically fine crystals of quartz, tridymite, and alkali feldspar. This explains why geologically ancient glasses are so rare and why glasses are most common in Tertiary and younger rocks. *See* IGNEOUS ROCKS; RHYOLITE

[C.A CA]

Volcano

An opening in the earth's crust through which magma (molten rock) or gases of magmatic origin, or both, issue. Only the products of volcanic activity are described here. For a discussion of processes of volcanism *see* VOLCANOLOGY, *see also* MAGMA; PETROLOGY

Types of volcanic vents. Fissures in the earth's crust, produced by orogenic (mountain-making) or other diastrophic forces, commonly constitute the channels through which magma rises toward the surface (*see* DIASTROPHISM). Magma consolidating in the fissure forms a dike, but that reaching the surface produces a fissure eruption. The fissures generally range in width from less than a foot to 10 feet, though many wider ones are known. During fissure eruptions of fluid lava, rows of fountains of liquid lava may extend for thousands of feet, or even several miles, along the fissure. The fissures may be independent of volcanic mountains or may occur on the flanks of large volcanic cones (Fig. 1), especially on shield volcanoes.

Fig. 1. Lava flow issuing from vents near base of Farallon de Pajaros, an active volcano in the Marianas Islands. (*Official U.S. Navy Photo*)

Other volcanic conduits are nearly cylindrical. Consolidation of material in these produces volcanic necks, or bosses. Typically, eruption of material from a cylindrical conduit builds a conical mountain terminating in a summit depression, or crater. The eruption may take place from fissure vents on the flanks of the cone (lateral eruption), but more commonly it occurs at the main, or central, vent (central eruption). Cones built by central eruptions often are arranged in lines, suggesting that they have been localized by fissures in the underlying rocks (Fig. 2).

Volcanic products. Molten rock extruded onto the earth's surface is called lava. It congeals into either volcanic glass, by quick chilling, or solid rock, by chilling and crystallization. The rock is also termed lava and sometimes lava rock. Depending on its viscosity during extrusion, molten rock may pile up steeply over the vent or pour out as streams known as lava flows. Lava torn into pieces and thrown into the air by explosions is called pyroclastic (fire-broken) material. *See* LAVA; VOLCANIC GLASS.

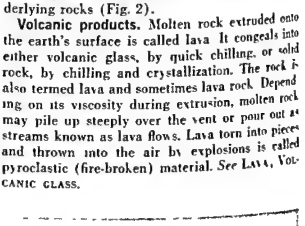


Fig. 2 Volcanic eruptions along fissure in earth's crust Island of Fayal, Azores. (*National Geographic Magazine*)

Birth of Kitsimbanyi volcano in Albert National Park, Belgian Congo (Top) Volcano crater split at center by lava flow (Bottom) River of molten lava, 300 yards wide, cutting through jungle (Photographs by R E Shanm)

The character of volcanic eruption depends ... and the general, ... viscous and contain less gas than more silicic (dacitic and similar) lavas Eruptions of mafic lavas commonly are quiet, with little explosion, and produce mostly lava flows; whereas eruptions of silicic lavas commonly are explosive and yield predominantly pyroclastic material. Some eruptions produce only pyroclastic material

Lava flows Several types of lava flows are recognized Flows with smooth or hummocky, gently undulating surfaces and with crust locally wrinkled into ropelike forms are called pahoehoe Ovoid mounds, a few feet high and a few tens of feet long, formed by buckling up of the crust, are known as tumuli Flows with very rough irregular surfaces covered with jagged spinose fragments resembling furnace clinker are called "aa" (Fig. 3). Flows in which the fragments comprising the upper part of the flow are fairly smooth-sided polygons are called block lava. Fluid mafic lavas characteristically form pahoehoe or aa flows, whereas less fluid lavas such as andesites more commonly form block lava flows. All three types commonly are vesicular, and may contain amygdules (*see* AMYGDULE)

Where pahoehoe flows pour into water they may form heaps of irregular ellipsoids, in cross section somewhat resembling so many sacks of grain or pillows These masses are known as pillow lava *See* PRECAMBRIAN

Emergence at the surface of highly gas charged silicic lava may be attended by a sudden frothing up as the contained gas comes rapidly out of solution The rapid expansion of the gas may tear the froth into an almost infinite number of tiny shreds, each of which quickly chills to volcanic glass. Continued ebullition of gas results in a mass of small solid or semisolid fragments, each surrounded by an envelope of still-expanding gas which pushes against all adjacent envelopes Thus the solid fragments are isolated from contact with each other, and the entire mass is given an expansive quality. The result is a very mobile suspension of solid fragments in gas These "froth flows" may advance with great speed down steep slopes and form voluminous deposits commonly known as welded tuffs or ignimbrites (*see* TUFF).

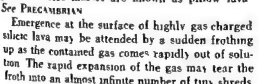


Fig 3 "Aa" lava flow advancing about 500 ft per hour, during the 1955 eruption of Kilauea volcano, Hawaii. The flow front is about 10 ft high. (G A Macdonald, U.S Geological Survey)

Similar expansive gas-solid suspensions may be formed by explosions, or explosion-induced collapses of active domes at the summit of volcanic cones, producing exceedingly mobile avalanches of hot gas-charged fragments that rush down the mountain side at speeds as great as 100 miles an hour, capped by a spectacular rolling, convoluting cloud of dust. Because of their incandescent nature, these have been called glowing avalanches, or *nuées ardentes*.

Pyroclastic materials. At depth, magma contains gas in solution, but as it rises into regions of lesser pressure near the surface of the earth, the gas starts to come out of solution and tends to escape from the liquid. From fluid lavas the gas escapes easily and quietly, with little explosion; but in more viscous liquids the gas may accumulate considerable pressure before it escapes and bursts forth in strong explosions Gradual crystallization of a magma beneath the surface may concentrate the gas in the residual fluid portion and, depending on the strength of the roof of the containing chamber, gas pressures may become very high and lead to violent explosions The explosions may throw into the air shreds of molten lava (magmatic, or essential, ejecta), fragments of already consolidated erupting lava (accessory ejecta), or solid fragments of older rocks (accidental ejecta)

Pyroclastic materials are classified also by their size, shape, and consistency Gobs or drops of material still liquid enough to assume rounded or drawn-out forms during flight are known as lapilli if they are 4–32 mm in average diameter and bombs if they are greater than 32 mm If they are still sufficiently plastic to flatten out when they strike the ground they are called Hawaiian, or pancake, bombs If they are sufficiently consolidated to retain their shape on impact, and acquired during flight a shape resembling a spindle or a summer squash, they are called spindle, or fusiform, bombs And if they are formed like a string or ribbon they are called ribbon bombs

Irregular fragments of frothy lava of bomb or lapilli size are called scoria or cinder; if the fragments are sufficiently plastic to flatten or splash as they hit they are called spatter The still molten surfaces of fragments of the latter type often adhere to each other to form welded spatter, or agglutinate Angular fragments larger than 32 mm either solid or too viscous to assume rounded forms during flight are known as *blocks* They may be essential, accessory, or accidental Accumulations of blocks are called breccia Ejecta smaller than ... ¼ mm ... ash or

Volcanic mudflows Mudflows are common on volcanic mountains where pyroclastic material is abundant. They may form by eruptions ejecting the water of a crater lake, or by breakdown of the confining wall of the lake, the water sweeping down the mountainside and forming a slurry of the loose material; or they may form by hot or cold avalanches descending into streams or onto snow or ice Probably by far the commonest cause, however, is simply heavy rain saturating a thick cover of loose unstable pyroclastic material on the steep slope of the cone, transforming the material into a mobile "mud" which rushes down the mountainside sometimes with a speed as great as 60 miles an hour, sweeping up everything loose in its path, and sometimes taking a heavy toll in property and human lives. Hot volcanic mudflows are sometimes called lahars

Somewhat related to mudflows are the great floods of water, known in Iceland as "Jokulhlaup," that result from rapid melting of ice by volcanic eruption beneath a glacier.

Fig 4 Paricutin volcano, Mexico, in eruption on May 24, 1943. Ash-laden steam and volcanic gas rising from cinder cone, which is about 1200 ft high (*U.S. National Museum*)

Volcanic landforms. Landscape features of volcanic origin may be either positive forms, the result of accumulation of volcanic materials, or negative forms, the result of lack of accumulation or of collapse

Positive forms. In certain areas voluminous extrusions of very fluid basaltic lava from dispersed fissure vents have built broad, nearly flat-topped accumulations covering areas as great as 100,000–200 000 square miles, with volumes of several tens of thousands of cubic miles These are known as *flood basalts*, or sometimes as *plateau basalts*. Two American examples are the Columbia and Snake river plains. Other examples are known in other parts of the world.

Where, instead of being dispersed, fissure eruptions of fluid lava have occurred repeatedly along the same zone of fissures, there results a broadly rounded dome-shaped hill or mountain known as a shield volcano. Some have volumes of several thousand cubic miles. The fissures extend from the summit down the flanks of the shield and are marked by long walls and rows of conelets of spatter Shield volcanoes consist almost wholly of innumerable superimposed thin lava flows

The eruption of less fluid lava, or lava with a higher gas content, produces more pyroclastic material. Cinder, bombs, and lapilli may accumulate around a central vent to form a conical hill or small mountain known as a cinder cone (Fig. 4) Alternations of lava flows and beds of pyroclastic material produce a composite volcano. Most of the well-known volcanic mountains are composite volcanoes; and some of them, such as Mayon in the Philippines, Fuji in Japan, and Shishaldin in the Aleutian Islands, are remarkably symmetrical cones of great beauty. Hills or small mountains composed predominantly or wholly of fine pyroclastic material are *ash cones*, or *tuff cones*

Lava too viscous to flow readily may pile up around and over the vent to form a steep-sided heap known as a volcanic dome. Slender spires thrust through apertures in such a dome are termed spines The famous spine formed at Mt Pelée, in Martinique, during the eruption of 1902, reached a height of nearly 1000 ft, but like most such spines was very short-lived.

Negative forms Small bowl-shaped depressions formed by explosion, or by failure of pyroclastic ejecta to accumulate directly above a vent, are known as craters. Most of them are found at the summit or on the flanks of volcanic cones, but some occur away from any cones

Larger depressions at the summit of volcanic cones are formed by collapse of the summit as the support beneath it is removed by the rapid withdrawal of magma, usually by surface eruption, but probably sometimes by migration of magma within the earth A depression of this sort is called a caldera. Probably the best-known example is that containing Crater Lake, in Oregon. Still larger, less regular depressions of similar origin are known as volcano-tectonic depressions. Like many calderas, their formation commonly, if not always, is associated with the eruption of great volumes of welded tuff

Submarine volcanism. In shallow water volcanic eruptions are very similar to those on land (Fig 5), though on the average probably somewhat more explosive owing to contact of hot lava with water and the resultant violent generation of steam. Ash of cones formed in this way may be rapidly altered by steam and hot water that saturates the cone during its formation, forming palagonitic tuff Such cones, like Diamond Head in Honolulu, may have broader, flatter profiles than those characteristic of cinder

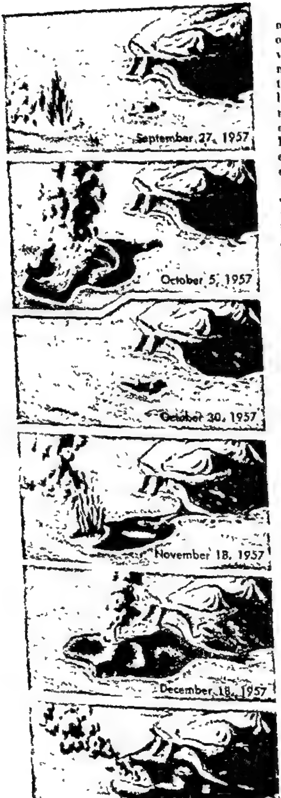


Fig. 5 Record of Ilha Nova's emergence from sea floor off island of Fayal in the Azores (*National Geographic Magazine*)

At great depths in the ocean the pressure of the overlying water may prevent the explosive escape of gas from erupting lava and greatly reduce the vesiculation of the lava itself. Pyroclastic materials may be absent in such an environment. Much of the Pacific Ocean appears to be floored by basaltic lava which, judging from its apparent density, is much less vesicular than the lavas of the basaltic cones that rise above it to form most of the Pacific Islands. This greater density may be the result of eruption of lava flows under the high pressures existing at the ocean bottom. *See* OCEANIC ISLANDS.

Fumaroles and hot springs. Vents at which volcanic gases issue without lava are known as fumaroles. They are found on active volcanoes during and between eruptions and on dormant volcanoes. They may persist long after the volcano has become extinct. The gases include water vapor, sulfur gases, hydrochloric and hydrofluoric acids, carbon dioxide and monoxide, and others in less abundance. They may transport and deposit at the surface small amounts of many of the metals. Temperatures of the escaping gases may reach 500–600°C. The halogen gases and metals generally are found in the high-temperature fumaroles. Lower-temperature fumaroles, in which the sulfur gases predominate along with steam, are called solfataras; and still cooler ones liberating predominantly carbon gases are called *mofettes*.

Fumaroles grade into hot springs and geysers (intermittent-spouting hot springs). The water of most, if not all, hot springs is predominantly of meteoric origin—rain water which has sunk into the rocks and moves through them, rather than water liberated from magma. Some hot springs appear to result simply from water circulating to warm regions at great depths in the earth's crust, but in many the heat is of volcanic origin and the water may contain volcanic gases. Indeed, the heat may be derived wholly from rising hot volcanic gases. The thermal springs and geysers of the Yellowstone region are of the latter sort. *See* THERMAL SPRING. [G.A.M.]

Bibliography: C. A. Cotton, *Volcanoes as Landscape Forms*, rev. ed., 1953. G. A. Macdonald and D. H. Hubbard, Volcanoes of Hawaii National Park, *Hawaii Nature Notes*, 4 (2), 1951. C. K. Wentworth and G. A. Macdonald, Structures and forms of basaltic rocks in Hawaii, *U.S. Geol. Survey Bull.* 994, 1953. H. Williams, Calderas and their origin, *Univ. Calif. Publ. Geol. Sciences*, 25 (6): 239–346, 1941. H. Williams, *The Geology of Crater Lake National Park, Oregon*, Carnegie Inst. Wash. Publ. 540, 1942.

Volcanology

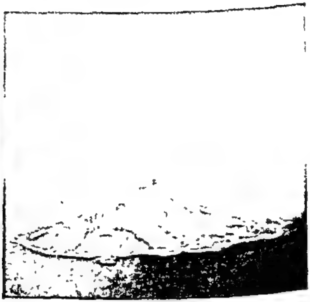
The scientific study of volcanic phenomena. Strictly speaking, it refers only to the surface eruption of magmas and related gases, and the structures, deposits, and other effects produced thereby. Commonly, however, it includes much of plutonic igneous geology, since effects at the earth's surface are the result of events at depth. For discussion of

(a)

(b)

(c)

(d)



(e)

(a) Fountain of liquid lava, 200 ft high, building a cone of cinder and spatter during 1955 eruption of Kilauea volcano, Hawaii (b) Small spatter cone, about 10 ft high, being built during the 1955 eruption of Kilauea (c) Mayon volcano, a symmetrical composite cone near Legaspi, Philippine Islands. (d) Didicas volcano, north [illegible] in eruption on June 17, 1952 Solid portions of dome protrude near top, lower slopes are mantled with talus formed by crumbling of the surface as the dome grew. (e) Farallon de Pajaros, Marianas Islands, in weak eruption on March 9, 1953 Lava flows have escaped near the base of the composite cone (a-c, G A Macdonald, USGS; d-e, Official U.S Navy Photo)

the surface structures and deposits resulting from volcanic activity, *see* VOLCANO. This article considers the surface activity of erupting volcanoes and the properties of erupting lavas.

Composition of magma. As formed at depth, magma consists of liquid rock containing dissolved gases. Rising toward the surface, it enters zones of lower temperature and pressure. Decreasing temperature tends to bring about crystallization, which produces solid crystals (*see* PHENOCRYST) suspended in the liquid. Other solid fragments (*see* XENOLITH) are incorporated from the walls and roof of the conduit through which the magma is rising. As crystallization progresses, volatiles and the more soluble silicate components are concentrated in the remaining liquid.

At some point in the rise of the magma, decreasing confining pressure and increasing concentration of volatiles in the residual liquid initiate the *separation of gas* from the liquid. From that point

dioxide and trioxide, elemental sulfur, carbon dioxide and *monoxide*, hydrogen, hydrochloric and *hydrofluoric acids*. All collections of volcanic gases

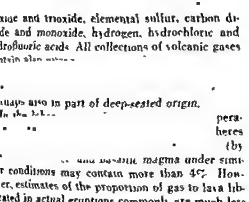
perhaps also in part of *deep-seated origin*.

pera-

heres

(by

... and basaltic *magma under* similar conditions may contain more than 4%. However, estimates of the proportion of gas to lava liberated in actual eruptions commonly are much less than that. Thus at the Hawaiian volcanoes, at Hekla in Iceland, at Nyamlagira in central Africa and at Vesuvius, volatiles constituted less than 1% of the weight of the lava *erupted* during the same interval. It is unlikely that such magmas are saturated with gas at *the higher* pressures existing even at depths *of only* a few kilometers in the earth. In *initially* unsaturated magmas, high gas pressures may be developed by the supersaturation *in* volatiles resulting from their concentration in the residual liquid phase during crystallization.



Cross-sectional diagram illustrating plutons and igneous activity in relation to geological structures and surface forms

Part of the gas liberated at volcanoes probably comes from the same deep-seated source as the silicate portion of the magma, but some is of shallower origin. Part of the steam is the result of superficial oxidation of deep-seated hydrogen, and some of the oxidation of the sulfur gases must have taken place close to the surface. At some volcanoes, such as Vesuvius, the carbon gases in part come from reaction of the magma with limestone at shallow depth. Ammonia and hydrocarbon gases present in some collections probably are derived from organic constituents of sedimentary rocks near the surface.

In some eruptions, such as the 1924 eruption of Kilauea, Hawaii, temperatures are low and the gas is wholly or very largely steam. Magmatic material may be entirely absent. In these phreatic eruptions the steam is simply heated ground water from the rocks adjacent to the volcanic conduit. *In other* eruptions, such as that of Paricutin, in Mexico, the large volume of steam *given off simultaneously* with lava and smaller amounts of magmatic gas far exceeds the theoretical saturation limit of the magma, and there is little doubt that it represents volatilized ground water.

Physical properties of magmas. The temperature of erupting magma has been measured in lava flows, and directly in volcanic vents, by means of thermocouples and optical pyrometers. Reasonably good and consistent measurements have been obtained for basic magmas. Thus at Hawaiian volcanoes the temperature of erupting lava in the vents ranges from a little above to a little below 1100°C, and that in moving lava flows ranges from about the same down to approximately 750°C. At Nyamlagira also, *in* vents and in flows close to vents, temperatures ranged from about 1040–1095°C. Temperatures during the 1950–1951 eruption of O-shima, Japan, were in the same range. Locally, temperatures as high as 1400°C may result from burning of volcanic gases *in* vents, and some secondary *heating* of lava flows may result from burning of hydrocarbon gases derived from vegetation buried by the flow.

Temperature measurements on more silicic lavas are much less accurate because of the greater violence of the *eruptions* and the *necessity* of working at considerable distances. *In general*, however, they suggest *lower* temperatures of eruption.

The viscosity of flowing basic lavas has been measured by *means* of penetrometers (*instruments* to measure the rate of penetration of the liquid by a slender *rod* under a given strength of thrust), and calculated from observed *rates* of flow in channels of known dimensions and slope. At Mauna Loa in Hawaii, lava close to the vents has a viscosity of 3000–4000 poises, increasing at greater distances from the vents until the flows *stiffen to immobility*. At Hekla, in Iceland, somewhat more silicic lava in the *vent has* a viscosity of 10,000 poises, and at Oshima, Japan, the lowest viscosi-

ties in two streams near the vent during the 1951 eruption were respectively 5600 and 18,000 poises In general, more silicic lavas are more viscous, though actual measurements of their viscosity are as yet lacking.

Types of volcanic eruptions. The character of a volcanic eruption is determined largely by the viscosity of the liquid phase of the erupting magma and the abundance and condition of the gas it contains Viscosity is in turn affected by such factors as the chemical composition and temperature of the liquid, the load of solid crystals and xenoliths it carries, the abundance of gas, and whether the gas is dissolved or separated as bubbles. In very fluid lavas small gas bubbles form gradually, and generally are able to rise through the liquid, coalescing to some extent to form larger bubbles, and escaping at the surface with only minor disturbance. In more viscous lavas the escape of gas is less easy, and produces minor explosions as the bubbles burst their way out of the liquid. In still more viscous lavas there appears at times to be a tendency for the formation of large numbers of small bubbles more or less simultaneously throughout a large volume of liquid, and the violent expansion of these may tear the frothy liquid to pieces and throw it into the air as a shower of volcanic ash and dust, accompanied by some larger blocks, or if gas is less abundant, it may produce an outpouring of frothy liquid as a froth flow. Also, gas may accumulate beneath a solid or highly viscous seal in the vent until it acquires enough pressure to burst its way forth in an explosion that hurls out fragments of the disrupted plug.

Types of eruptions customarily are designated by the name of a volcano or volcanic area that is characterized by that sort of activity (though probably all volcanoes show somewhat different sorts of activity at different times).

Hawaiian eruptions. Eruptions of the most fluid lava, in which relatively small amounts of gas escape freely with little explosion, are given this designation Most of the lava is extruded as thin flows that spread to distances of several miles from their vents. Clots of lava thrown into the air remain fluid enough to flatten out on striking the ground, and commonly to weld themselves together. An occasional feature of Hawaiian activity is the lava lake —a pool of liquid lava with convectional circulation that appears to represent simply the top of a column of magma occupying the volcanic conduit

Strombolian eruptions. These somewhat more explosive eruptions of lava, with greater viscosity, produce a larger proportion of pyroclastic material. Many of the bombs and lapilli assume rounded or drawn-out forms during flight, but are sufficiently solid to retain these shapes on impact.

Vulcanian eruptions. Generally still more explosive are the Vulcanian type. Angular blocks are hurled out of viscous or solid lava, commonly accompanied by voluminous clouds of ash but with little or no lava flow.

Pelean eruptions. Characterized by the heaping up of viscous lava over and around the vent, these extrusions form a steep-sided hill or volcanic dome. Collapses of portions of the dome may result in glowing avalanches (*nuées ardentes*).

Plinian eruptions. These paroxysmal eruptions of great violence are characterized by voluminous explosive ejections of pumice, and by froth flows. The copious extrusion of siliceous magma commonly is accompanied by collapse of the top of the volcanic cone, forming a caldera, or of a broader region, forming a volcano-tectonic depression The Roman naturalist, Pliny, was killed in 79 A.D. by such an eruption of Vesuvius.

Ultravulcanian explosions. In contrast to the foregoing magmatic eruptions, some low-temperature explosions of this type throw out fragments of old volcanic or nonvolcanic rocks, accompanied by little or no magmatic material. Certain explosion pipes and pits, known as diatremes and maars, have been produced by ultravulcanian explosions.

Phreatic explosions. Volatilization of ground water when it comes in contact with hot lava or hot rocks in the walls of the volcanic conduit causes these disturbances; occasionally they are of considerable violence

Surface deformations from eruptions. Measurements at certain volcanoes have demonstrated that the entire volcanic mountain swells and shrinks, apparently in response to changes in conditions beneath the volcano. The measurements are of two sorts: precise leveling surveys in which the change of altitude of a series of benchmarks is related to some point outside the disturbed area, and measurement of the change in inclination of a given point on the slope of the cone by means of instruments known as tiltmeters. Some tiltmeters are capable of detecting changes in inclination measured in fractions of a second of arc, but the total changes of level resulting from tumescence and detumescence of the volcano may amount to several feet Thus, between 1912 and 1921 a benchmark at the edge of Kilauea caldera, Hawaii, rose about 3 ft and between 1921 and 1926 the same benchmark sank 3½ ft. During the latter interval a benchmark at the edge of the inner crater, Halemaumau, sank 6 ft

Even more extreme changes have occurred elsewhere During the 1944–1945 eruption of Usu, Japan, the ground surface near the foot of the mountain was upraised in a broad dome 50 m high before explosions finally burst through its top and a protrusion of viscous lava took place In this instance it is clear that the deformation resulted from an intrusion of lava beneath the uplifted area. At the Hawaiian volcanoes the mechanism is less clear, and is generally referred to simply as a change in volcanic pressure

At least in some instances the swelling and shrinking of Hawaiian volcanoes show direct close relationship to eruptions, the mountain swelling before eruption and shrinking during or immedi-

ately after it. This suggests swelling as a result of injection of magma into some sort of reservoir in or beneath the cone, and shrinking as a result of drainage of the reservoir.

Similar subsidences have accompanied eruptions in other regions. Thus, during the 1914 eruption of Sakurajima, Japan, an area 47 km across, centered at the volcano, sank more than 15 cm, and the sinking in the central part of the area exceeded 80 cm The great collapses that form calderas and volcano-tectonic ...

... gether with tumescence of the volcano, constitute the principal evidences of coming eruption For the most part the ...

... monic tremors, which may go on continuously with little or no change for *many hours, days, or weeks;* and spasmodic tremor, which is simply a closely spaced succession of minute earthquakes (*see* ...ISMOLOGY) Volcanic explosions produce shock ...ves in the atmosphere—sometimes seen as flash...; arcs rendered visible by refraction of light in ... denser wave front Explosions or collapses tak...; place under water may cause water waves, ...netimes of great size These fall within the class tsunamis (seismic sea waves) though most tsu...mis have other causes Waves of this sort accom...nied the collapse of Krakatoa volcano, in 1883 ...d destroyed many villages on the shores of neigh...ring Java and Sumatra. *See* TSUNAMI [G A M]

Bibliography R A Daly, *Igneous Rocks and ... Depths of the Earth*, 2d ed., 1933; J Gilluly, ... C. Waters, and A O. Woodford *Principles of ...ology*, 1957; T A Jaggar, *Origin and Development of Craters*, Geol Soc Am Mem. 21 1947, ...A Macdonald Activity of Hawaiian volcanoes ...ring the years 1940–1950 *Bull volcanol.*, vol 15 ...54, G A Macdonald *Hawaiian Volcanoes dur... 1952*, USGS Bull 1021 B. 1955

...olt

...e unit of potential difference or electromotive ...ce in the meter kilogram second (mks) system ... electrical units By definition two points are at a ...tential difference of 1 volt if 1 coulomb of elec...city will do 1 *joule* of work in going from one ...int to the other *See* ELECTRICAL STANDARDS, ...ECTRICAL UNITS, ELECTROMOTIVE FORCE (EMF), ...OTENTIAL DIFFERENCE [R P WI]

...oltage amplifier

... electronic amplifier circuit that produces an ...tput of greater voltage magnitude than the input ...ltage For a general discussion of amplifiers, *see* ...MPLIFIER.

Voltage amplifiers are built to amplify signals ...th frequency components in the range from zero ...cles per second (cps) to thousands of megacycles (Mc). To cover this frequency range several different types of amplifiers are required. In the range from zero to about 50 kilocycles (kc) direct-coupled amplifiers may be used. In the range from a few cycles per second to about 30 kc, amplifiers are classified as audio. The upper range is often extended to about 10 Mc for use in instruments, and the amplifiers are classified as video. For the frequency range from a few hundred kilocycles and up, tuned amplifiers are normally used. At ultrahigh frequencies, amplifiers designed around special tubes are employed (*see* KLYSTRON; LIGHTHOUSE TUBE; TRAVELING-WAVE TUBE).

RC-coupled amplifier. The *RC*-coupled amplifier is employed in the frequency range from a few cycles per second to about 10 Mc. However, operation at the higher frequencies in this range is obtained only by employing frequency-compensation networks. The term *RC*-coupled amplifier generally refers to any set of amplifier stages that are coupled by a resistance-capacitance network. Figure 1*a* illustrates a grounded-cathode circuit. The coupling capacitor *C* is required to block the dc voltage at the plate of the first tube from ap-

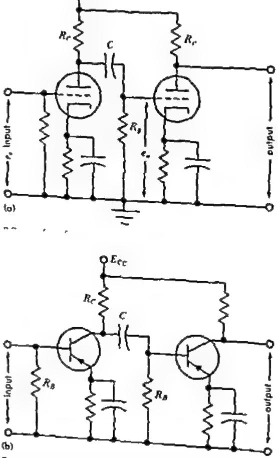


Fig 1 *RC*-coupled amplifiers (*a*) Vacuum tube (*b*) Transistor E_{bb} is the plate supply voltage and E_{cc} is the collector supply voltage

pearing at the grid of the following tube. This capacitor limits the lower frequency for which the amplifier will have a usable gain, because the reactance $1/2\pi fC$ of the capacitor is inversely proportional to frequency. Therefore, as the frequency decreases, the amplified signal voltage at the plate of the tube will be attenuated by the voltage divider formed by this reactance and the grid resistor.

A similar condition occurs when two transistor stages are *RC*-coupled as in Fig. 1*b*. Here, two common-emitter stages are coupled by means of capacitor *C*, and the same problem of reduced gain at lower frequencies exists.

The high-frequency limit of usable amplification is determined by the shunt capacitance and the effective plate load resistance of a tube, or output load resistance of a transistor. The shunt capacitance is composed of the input capacitance of the following tube or transistor, the plate-to-cathode capacitance of the tube or output capacitance of the transistor in the stage, and stray wiring capacitance introduced by the placement and connection of the components.

The input capacitance to a triode tube in a grounded-cathode amplifier stage is composed of the capacitance formed from the grid-to-cathode and from the plate-to-grid capacitance multiplied by 1 minus the gain of the stage. The gain is negative in the midband region, permitting the input capacitance to be large, on the order of 30–50 micromicrofarads ($\mu\mu f$).

If the upper and lower half-power (or cut-off) frequencies are respectively high and low, compared with the frequency spectrum of the signal to be amplified, then the gain is essentially uniform with frequency.

Mid-frequency gain. In the case of the vacuum-tube amplifier the mid-frequency gain is

$$A_{\text{mid}} = -g_m R_{eq}$$

where g_m is the transconductance of the tube, the minus sign indicates a 180° phase reversal between the input grid signal and the output signal, and the equivalent load resistance R_{eq} is the parallel combination of the dynamic plate resistance r_p of the tube and the load or coupling resistance R_c.

For the transistor amplifier, the computation of gain is more involved, because of the relatively lower input and output impedances of the transistors. The mid-frequency voltage gain is

$$A_{\text{mid}} = A_i \frac{R_L}{R_{in}}$$

where the input resistance of this stage is

$$R_{in} = \frac{\Delta^h R_L + h_{11}}{h_{22}R_L + 1}$$

and R_L, the total load resistance seen by this transistor, includes the parallel combination of the coupling resistor R_C, the bias resistor R_B, and the input resistance R'_{in} of the next stage. The h_{11} is the input impedance, h_{22} the output admittance of the transistor, h_{12} the voltage-feedback ratio, and h_{21} the forward-current ratio, and the quantity Δ^h is $h_{11}h_{22} - h_{21}h_{12}$. Because of the dependence of the gain on the R_{in} of the next stage, it is usually necessary to compute the gain per stage in a transistor amplifier by starting with the last stage and working backward toward the input. The current gain is

$$A_i = \frac{h_{21}}{h_{22}R_L + 1}$$

High-frequency gain. The upper half-power frequency is that frequency for which the reactance of the shunt capacitance is equal to the equivalent load resistance.

For the vacuum-tube amplifier the equivalent load resistance R_{eq} is the parallel combination of r_p, R_C, and R_g, the grid resistance in Fig. 1*a*. The upper half-power frequency is

$$f_2 = \frac{1}{2\pi R_{eq}C_{\text{shunt}}}$$

For the transistor amplifier the equivalent resistance is the parallel combination of R_C, R_B, the input resistance of the next stage R_{in}, and the output resistance of the transistor R_o. R_o is given by

$$R_o = \frac{h_{11} + R_g}{\Delta^h + h_{22}R_g}$$

The upper half-power frequency is formed from the same equation as given for the vacuum-tube amplifier, differing only in the quantity R_{eq}.

Low-frequency gain. The lower half-power frequency is that frequency for which the reactance of the coupling capacitor *C* equals the effective resistance *R* appearing in parallel with the coupling capacitor. For the vacuum-tube amplifier at low frequencies, the effective resistance is the dynamic plate resistance r_p and the coupling resistance R_C in parallel, and this combination in series with the grid resistor R_g of the next stage. Therefore the lower half-power frequency is

$$f_1 = \frac{1}{2\pi RC}$$

In the case of the transistor amplifier, the resistors appearing across *C* are the output resistance R_o, in parallel with the coupling resistor R_C, and this combination in series with the parallel combination of the bias resistor R_B and the input resistance R'_{in} of the next stage. The same expression for f_1 is applicable.

Pass band. At the half-power frequencies the gain is equal to the midband gain divided by the square root of 2. For either the vacuum tube or transistor *RC*-coupled amplifier the gain at any frequency *f* can be expressed in terms of the midfrequency gain. At low frequencies

$$A_{\text{low}} = \frac{A_{\text{mid}}}{1 - j(f_1/f)}$$

and at high frequencies

$$A_{\text{high}} = \frac{A_{\text{mid}}}{1 + j(f/f_2)}$$

This simple relationship permits drawing a universal frequency-response curve *A*, Fig 2, applicable to any single stage of either an electron-tube or transistor amplifier. The graph is made universal by plotting relative gain as a function of the ratio of the signal frequency to the half-power-point frequencies. The slope of the curve approaches −6 decibels (db) per octave for a single stage of amplification One octave is a change in frequency by a factor of 2. The frequency range $f_2 - f_1$ is called the bandwidth *B* of the stage

The gain of a multistage amplifier is equal to the product of the gain of each of the individual stages. Curve B of Fig. 2 is for three identical stages For this unique case the slope of the curve is −18 db per octave, which is the sum of the −6 db per octave slope inherent in each of the three stages. When stages are cascaded, the half-power (70.7%, or −3 db) points are moved in toward the mid frequencies. Therefore, the cascading of stages decreases the frequency pass-band between the half-power points

For reasons of economy, amplifiers used for entertainment purposes and similar applications sometimes make use of the frequency regions beyond the flat mid frequency ...

...sumed to the mid frequency range

Interstage transfer function One other important distinction arises between the vacuum tube and the transistor The vacuum tube is inherently a *voltage* amplifier, and a voltage appearing across the grid resistor R_g in Fig 1*a* also appears from grid to cathode in the next tube and is amplified The situation is considerably different with transistors, which are current amplifiers A signal current leaving the first transistor collector follows two paths, each of which shunts a portion of this current to ground. This reduces the current gain and must be accounted for in computing the over-all amplifier current gain

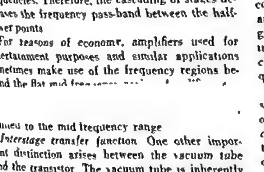


Fig 2 Universal frequency-response curve for *RC*-coupled amplifiers Curve A is for any single stage of amplification Curve B is for three identical stages

The interstage transfer function is given by the ratio of the current flowing into the input resistance R'_{in} of the next stage, divided by the current out from the preceding transistor. In terms of the parameters of the circuit, the transfer function is

$$a = \frac{R_C R_B}{R_C R_B + R'_{in}(R_C + R_B)}$$

For well-designed, broad-band amplifiers this transfer function runs in the neighborhood of 0.7.

The over-all voltage gain of a vacuum-tube amplifier is

$$A_{\text{total}} = A_1 \cdot A_2 \cdot A_3 \cdots$$

and for a transistor amplifier the over-all current gain is

$$A_{\text{total}} = a_1 \cdot A_1 \cdot a_2 \cdot A_2 \cdot a_3 \cdot A_3 \cdots$$

where A_1, A_2, ..., are the gains per stage, and a_1, a_2, . . . represent the interstage transfer functions.

Impedance-coupled amplifier. The impedance-coupled amplifier is formed from the *RC*-coupled amplifier by replacing the plate load resistor or the grid-leak resistance, or both, with audio chokes (an iron-core inductor having a large inductance) The circuit is not in common use because of poor frequency response

Transformer-coupled amplifier. This amplifier, shown in Fig. 3, may be either an audio amplifier or a tuned amplifier operating in the radio-frequency region. The voltage gain of a multistage transformer coupled amplifier can usually be made considerably greater than that obtainable with the same stages coupled by *RC* networks.

Audio amplifier. In the early stages of development of the audio amplifier, transformers were used for coupling between stages The gain per stage could be increased by employing an iron-core transformer with a secondary-to-primary voltage ratio *n* greater than unity This increase in voltage gain is obtained at the expense of the frequency response of the amplifier

For amplifying frequencies below about 20 kc, the transformer-coupled amplifier may be more economical and possibly of smaller physical size than a comparable *RC*-coupled amplifier. The choice is a problem in economics

For low frequency vacuum-tube transformer-coupled amplifiers the maximum usable gain is

$$A = \mu n$$

where μ is the amplification factor of the tube and *n* is the secondary-to-primary turns ratio of the transformer. Typical turns ratios are 3.5:1, 2.5:1, and 1:1 for interstage transformers. In general, the frequency response varies inversely with turns ratio Therefore, increasing the turns ratio to obtain voltage gain results in a reduction in available frequency response

When transistors are employed, the principal function of the transformer is not to supply additional voltage gain but rather to provide an im-

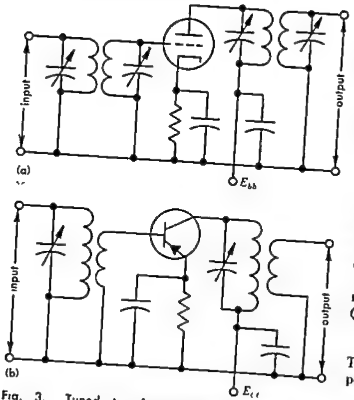


Fig. 3. Tuned transformer-coupled amplifier. (a) Vacuum tube (b) Transistor. The plate supply voltage is E_{bb} and the collector supply voltage is E_{CC}.

pedance match between the output resistance of one transistor and the input resistance of the next. If proper impedance matching is achieved, the maximum available gain (MAG) of each transistor is theoretically obtainable. The MAG is determined solely by the parameters of the transistor, in much the same manner that the amplification factor μ is a parameter of the vacuum tube alone. The MAG is given by

$$MAG = \frac{h_{21}^2}{[(\Delta^h)^{1/2} + (h_{11}h_{22})^{1/2}]^2}$$

where the h quantities have been previously defined. The appropriate h quantities for the transistor configuration employed must be used. *See* TRANSISTOR CONNECTION.

Radio-frequency amplifier. Transformer-coupled amplifiers are used for rf and i-f voltage amplifiers. These are usually double-tuned transformers (the primary and secondary are both tuned with suitable capacitors) operating over a frequency bandwidth that is small compared to the resonant frequency. The transformers usually have either air or powdered-metal cores. The core slug is about as long as the length on the coil form for the windings. Tuning is sometimes accomplished by moving the core partially in and out of the coil. *See* TRANSFORMER.

The gain of an rf double-tuned transformer-coupled amplifier (including i-f amplifiers) is determined to a great extent by the efficiency of the resonant circuits (Fig. 3). The efficiency of the inductor in any tuned circuit can be described in terms of a factor of merit

$$Q = \frac{\omega L}{R}$$

where L is the inductance in henries, R is the effective series resistance of the coil and ω equals 2π times the frequency at which Q is evaluated.

For the double-tuned transformer-coupled vacuum-tube amplifier, the gain at the resonant frequency f_0 is

$$A = g_m k \frac{\omega_0\sqrt{L_pL_s}}{k^2 + 1/Q_pQ_s}$$

The subscripts indicate primary and secondary inductances and Q's. The transconductance of the tube is g_m, $\omega_0 = 2\pi f_0$, and k is the coefficient of coupling between the coils.

For critical coupling ($k = 1/Q_pQ_s$) and nearly identical primary L_p and secondary L_s coils ($Q_p = Q_s = Q$) the voltage gain at resonance is simply

$$A = \tfrac{1}{2} g_m\omega_0 LQ$$

The associated bandwidth between half-power points is

$$B = \frac{\sqrt{2}f_0}{Q} \quad \text{cps}$$

when the coupling is critical.

When transistors are employed, the possibility of obtaining considerable voltage gain from the tuned transformer, as is possible with vacuum tubes, becomes more difficult because of the relatively low input and output impedances of transistors. The tuned transformer functions more as an impedance-matching device. Taps or auxiliary windings are often provided on rf and i-f transformers for use with transistors to provide means for impedance matching.

Cathode-follower stages. If the plate of the vacuum tube is connected directly to the power supply source and the output is taken from the cathode the resulting circuit has special properties which are useful.

The vacuum tube may be replaced with a grounded-collector transistor configuration and the

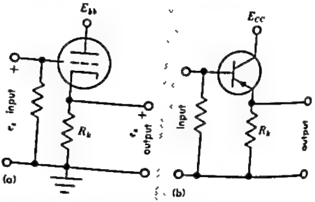


Fig. 4. (a) Cathode follower amplifier (b) Grounded collector amplifier. The plate supply voltage is E_{bb} and the collector supply vol

same over all results are obtained. A simple version is illustrated in Fig 4 The voltage gain of the circuit is less than unity, typical values ranging from 0.90 to 0.95 The input and output voltages are in phase. The principal properties of interest are the relatively high input impedance and the low output impedance. The amplifier is often used as an impedance-matching device. In such a situation a cathode follower can be inserted between a source and a load with only a slight decrease in voltage gain Current and power gain may be quite high if desired

The input capacitance of a cathode-follower stage is low, since the gain is very nearly equal to unity and is positive Therefore, the input capacitance is essentially the grid-to-cathode capacitance, which is on the order of 2 $\mu\mu f$. Thus the cathode follower is ideally suited for use in the interstage coupling of a broad band amplifier

The mid frequency gain of a cathode follower circuit is given by

$$A = \frac{\mu R_k}{r_p + (\mu + 1)R_k}$$

where μ is the amplification factor, r_p is the dynamic plate resistance, and R_k is the cathode load resistor The output impedance that the cathode follower presents to its output circuit is of interest and is given by

$$Z_{out} = \frac{R_k}{1 + g_m R_k + R_k/r_p}$$

where g_m is the transconductance By a proper selection of the tube and the cathode load resistor R_k, a wide range of output impedances may be obtained

If transistors are used, then the analysis proceeds as previously outlined for a grounded emitter

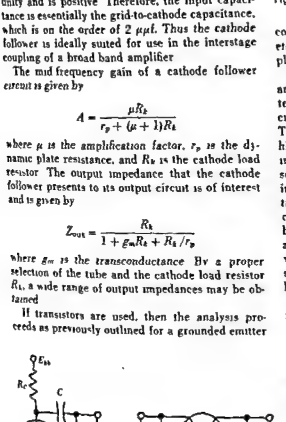


Fig 5 (a) Grounded-grid amplifier (b) Grounded-base amplifier E_{bb} is the plate supply voltage, E_{EE} the emitter supply voltage, and E_{CC} the collector supply voltage

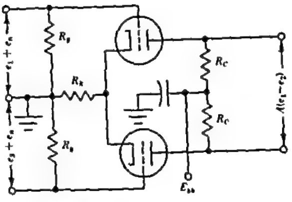


Fig 6. Differential amplifier.

configuration except that the appropriate h parameters for grounded-collector operation are employed in the calculations.

Grounded-grid amplifier. If the grid is grounded and the input signal is applied to the cathode terminal, with the output taken from the plate, the circuit is called a grounded-grid amplifier (Fig 5). This circuit features a low input impedance and a high output impedance Because of the low input impedance the amplifier will draw power from the source and must, therefore, be used with a low-impedance source It has found its chief application in rf amplifiers operating at very high frequencies, such as the tuner in a television receiver, because at these frequencies it is capable of giving a better signal-to-noise ratio than that obtainable with a pentode tube It does so at the expense of rf selectivity This excludes it from use in a narrow-band receiver application.

The voltage gain of the grounded grid amplifier is given by the equation

$$A = \frac{(\mu + 1)Z_L}{r_p + (\mu + 1)R_s + Z_L}$$

where R_s is the internal resistance of the signal source, μ and r_p are the amplification factor and ... tube, ... id-fre- ... is the ... or R_L and the grid resistor of the next stage R_g

The input impedance of the stage is

$$Z_{in} = \frac{r_p + Z_L}{\mu + 1}$$

The stage can be used to match low-impedanc circuits to high-impedance circuits. The corr sponding transistor circuit would be a grounde base configuration, and the appropriate h parar eters for this arrangement would be employed the calculations as outlined for the grounde emitter configuration.

Direct-coupled amplifier. When the coupli between successive stages of an amplifier is su that direct current may flow from the output of o

stage to the input of the following stage, the coupling is said to be direct.

A common form of direct coupling between two stages is found in the cathode-coupled amplifier. Here a cathode follower is direct-coupled to a grounded-grid amplifier. The circuit finds application as a dc and low-frequency amplifier.

Two or more grounded-cathode or grounded-emitter stages may be direct-coupled by operating the ground of each successive stage at the quiescent (average or dc) potential of the plate or collector of the previous stage. The major disadvantage to this scheme is the large supply voltages required. For further details, *see* DIRECT-COUPLED AMPLIFIER.

Differential amplifier. Also called a difference amplifier, this is a modification of the cathode-coupled amplifier. The two signals for which the difference is required are coupled through a common cathode resistor R_k (Fig 6). Dual inputs are shown with applied signal voltages e_1 and e_2. Also an equivalent input noise signal e_n is indicated at each input. This noise voltage represents all the noise arising from changes in tube parameters, emission, stray pick-up, and so forth (*see* NOISE, ELECTRICAL). If precautions are taken to balance the two circuits, the noise voltages will cancel and the output will be simply the difference of the input signals.

A low-noise, single-signal amplifier is made by grounding one of the grids and applying the signal to the other. The noise voltage still is present at each grid and cancels in the output, making possible low-noise amplification of the signal.

[H F.K.]

Bibliography: R. L. Riddle and M P Ristenbatt, *Transistor Physics and Circuits*, 1958; J. D. Ryder, *Engineering Electronics*, 1957; S. Seely, *Electron Tube Circuits*, 2d ed., 1958; F E. Terman, *Electronic and Radio Engineering*, 4th ed., 1955.

Voltage measurement

The determination of potential difference between two points in an electric circuit. Voltage is usually measured by electromechanical instruments having capacity, input impedance, and other characteristics suited to the kind of voltage and kind of circuit under test (*see* POTENTIAL, ELECTRIC; VOLT). The scope of voltage measurements ranges from a few microvolts to hundreds of kilovolts. The frequency varies from 0 (direct current) to about 2000 megacycles per second (Mc). Maximum errors may range from 5% to 0.05% of the reading, depending on the need for accuracy and the instrument used.

An indicating voltmeter indicates voltage by movement of a pointer over a calibrated scale. A recording voltmeter makes a graphic record of the voltage by movement of a pen on a calibrated chart. *See* RECORDING INSTRUMENTS, GRAPHIC; VOLTMETER.

Calibration. Industrial and laboratory voltmeters are usually secondary instruments with arbitrary calibrations in units of voltage or some multiple thereof. Direct-current voltmeters are calibrated by comparison with a standard-voltage cell, using potentiometers. The calibration of ac voltmeters is also based on the standard cell, using potentiometers and a dc-ac transfer device, such as a precision electrodynamometer or a thermal converter.

The primary standard of electromotive force (emf) for the United States is a group of Weston saturated cadmium cells maintained at constant temperature at the National Bureau of Standards (*see* WET CELL). The saturated cadmium cell employs mercury as a positive electrode and cadmium in a mercury amalgam as a negative electrode, with saturated cadmium sulfate solution as the electrolyte. The cell has an emf of 1.01830 volts at 20°C. The unsaturated cadmium cell, in which the electrolyte is unsaturated at ordinary temperatures, is a more suitable working standard, having a lower temperature coefficient and greater portability than the saturated cell. This form of standard cell is almost universally used in standardizing laboratory work. The emf of the unsaturated cell is 1.0190 volts when new. It decreases gradually with time and should be discarded when the emf falls to 1.0183 volts. For circuits used with these cells for the standardization of voltmeters *see* POTENTIOMETER (VOLTAGE MEASUREMENT).

Methods. Voltages in ac power circuits are usually sine waves, and the voltage measurement may require determination of average, root-mean square (rms), or peak values of this wave. In general, the rms, or effective, value is desired, so the instrument is designed to square and integrate the instantaneous value, thus

$$V = \frac{1}{T}\sqrt{\int_0^T v^2\,dt} \qquad (1)$$

where V is the instrument reading, T is the period of the wave, v is the instantaneous voltage, and t is time.

Average-reading voltmeters are those which integrate the first power of the instantaneous voltage, their response being otherwise the same as rms voltmeters. Averaging is a nearly universal characteristic of rectifier-type voltmeters, which may, however, be calibrated for rms measurement of sine-wave voltages. This calibration is valid only for sine waves, and errors are introduced if the wave is distorted. *See* ELECTRICAL MEASUREMENTS.

In certain high-voltage tests for dielectric strength the peak, rather than average or rms, value is required. Here, the spark-gap method of voltage determination is used. Two conducting spheres, separated a distance corresponding to the required voltage, are connected to the source and the voltage raised until arc-over occurs. The peak voltage for arc-over is determined either by calculations based on the dielectric strength of air, or by separate calibration of the sphere-gap separation with a voltmeter and potential transformer.

Since all voltage measurement requires transfer of energy from the circuit to the measuring device, it is necessary to ensure that this transfer does not seriously disturb the voltage [illegible]easurement.

In industrial measurements the power required by the voltmeter is generally insignificant relative to the capacity of the source. In communications and electronics, measurements are required on low-energy sources, and instruments with impedances in the megohm range may be required. Correction for instrument losses, where necessary, are made by the usual methods of circuit computation. [A.J.C.]

Measurement of small voltages. Potentiometer measurements of dc voltages can be made down to 1 microvolt (μv) or less, but great care is required in minimizing the effects of thermal emfs. More convenient and rugged instruments of several types are available for direct, accurate indication of small dc voltages, such as those from thermocouples. These include systems in which the direct current is changed to alternating current by a modulating device, and the alternating current is then amplified and rectified to operate a meter. Suitable modulators include choppers of the vibrating-contact and photoconductive types, vibrating capacitors, and second-harmonic magnetic modulators.

Several galvanometer amplifying systems are available in which a moving-coil element deflects a light beam onto a balanced pair of photocells, giving a large output for minute movement of the coil. In a similar system the moving element modulates the frequency of an oscillator, changing the output of a frequency-sensitive circuit. Galvanometer amplifying systems and chopper-modulated systems can be used with full-scale sensitivities down to the order of 1 μv. Magnetic amplifiers are satisfactory for voltages of 1 millivolt (mv) or less in low-resistance circuits. The bandwidth of all these devices is limited to several cycles per second for the more sensitive designs, and to about 30 cps for a range of 10 mv.

Straight dc amplifiers using vacuum tubes are employed when wider bandwidth is required or when very little current can be drawn from the source, but these are limited to sensitivities of the order of 30 mv full scale. When special electrometer tubes are used, dc amplifiers [illegible] input resis[illegible] can be use[illegible] fier withou[illegible]ing its bandwidth.

Alternating voltages as low as 0.1 mv at frequencies up to about 5 Mc can be measured conveniently by amplifier-type vacuum-tube voltmeters. Voltages of less than 1 μv can be measured by potentiometers at audio frequencies and by standard signal generators, using a substitution method, over the entire radio-frequency spectrum.

Nonsinusoidal voltage measurement. In some cases the rms value of a complex waveform is the quantity desired. In other cases the fundamental component is to be measured in the presence of harmonics, or the various harmonics are to be measured separately. Thermal devices, such as thermocouples and thermistors, are convenient for rms measurements. Oth[illegible] [illegible]nts having a square-law relationship between response and voltage can also be used. Electrodynamometer instruments respond accurately to rms values within their frequency range, and moving-iron instruments do so to a close approximation. Rectifier instruments give a square-law response at low voltages, so are useful over a limited range.

A wave analyzer, consisting of an amplifier, a sharply tuned filter, and an indicating meter, can be used to measure the harmonic components separately. If only the fundamental is to be measured, a phase-sensitive detector, such as a ring modulator, can be used. This must be excited by a voltage of the fundamental frequency which is matched in phase to the component to be measured. *See* HARMONIC ANALYZER.

High-frequency voltage measurement. Commercial thermocouple instruments are primarily for current measurement, and are usually limited to audio frequencies when used with a series-resistance multiplier for voltage measurement. They can be used as low-voltage calibrating elements, however, at frequencies up to about 100 Mc. Vacuum-tube diode-rectifier voltmeters can be used at frequencies beyond 1000 Mc. Crystal diode rectifiers can be used at somewhat higher frequencies but can handle only a few volts. Meaningful voltage measurements beyond about 100 Mc depend on the measuring element not disturbing the circuit, or on the ability to relate the voltage measured to a known point in the circuit. Hence the measuring element may be permanently wired into the apparatus, or a carefully designed coaxial-line fixture can be used between the source and the load to house the rectifier and prevent energy reflections. At frequencies beyond about 2000 Mc, voltages are not usually measured directly but are calculated from measurements of transmitted power, made with a directional coupler, and impedance measurements. *See* VACUUM-TUBE VOLTMETER.

Oscilloscope methods. The more elaborate commercial oscilloscopes are available with a coordinate scale and a source of calibrating voltage. The amplitudes of ac waveforms can be measured with an accuracy of a few per cent and a sensitivity depending on the transmitted bandwidth. A typical instrument gives 50 μv/cm deflection with an amplifier bandwidth of 60 kc, or 50 mv/cm with a bandwidth of 20 Mc. Oscilloscopes with direct-coupled amplifiers can be used to measure composite voltages having dc and ac components. *See* CURRENT MEASUREMENT, DIGITAL VOLTMETER; OSCILLOSCOPE, CATHODE-RAY. [W.N.T.]

Voltage regulation

The rise in voltage of a generator when rated full-load conditions are removed, under constant excitation and speed, divided by normal rated voltage.

$$\begin{pmatrix}\text{Voltage}\\ \text{regulation}\\ \text{in per cent}\end{pmatrix} = \frac{\begin{pmatrix}\text{no-load}\\ \text{voltage}\end{pmatrix} - \begin{pmatrix}\text{rated full-}\\ \text{load voltage}\end{pmatrix}}{\text{rated full load voltage}} \times 100$$

The voltage regulation of a generator is an indica-

tion of how sensitive its voltage will be to load changes under manual excitation control. A low value is desirable to give a constant voltage output. However, in large generators, a high value is often desirable, since it limits short-circuit currents and therefore protects the machine from damage.

[L.T.R.]

Voltage regulator

A device or circuit that maintains a voltage nearly constant over a range of variations of input voltage and load current. Voltage regulators are used wherever the unregulated voltage would vary more than that which can be tolerated by the electrical equipment using that voltage. Alternating-current distribution feeders use voltage regulators to keep the voltage supplied to the user within a prescribed range. Electronic equipment often has voltage regulators in the dc power supplies.

ELECTRONIC VOLTAGE REGULATORS

In much electronic equipment the dc power-supply voltage must remain constant in spite of input ac line voltage variations and output load variations. To provide a constant output voltage, two types of voltage regulator circuits are used in electronic power supplies, the gas-tube regulator circuit and the vacuum-tube regulator circuit. The voltage regulator circuit may also augment the filtering action of the power-supply filters, and it raises or lowers the internal impedance of the power supply as seen from the load. The price paid for the regulator action is a considerable drop in voltage, thus considerable power is lost in the regulator circuit. Most of the power is lost either in a regulator resistor or in the regulator tubes.

Gas-tube regulator circuit. One gas-tube regulator circuit is shown in Fig. 1. The tube used is a glow tube, which is characterized by almost a constant voltage over the specified range of tube current, as shown in Fig. 2. The constant voltage depends upon the particular type of gas tube employed and may range from about 60 to 150 volts.

The circuit of Fig. 1 depends on two circuit conditions, (1) the input voltage E_i is always the sum of the voltage drop V_s across the series resistor R_s and the load voltage V_L which is maintained constant by the VR tube, and (2) the current through the series resistor is always the sum of the load current i_2 and the VR-tube current i_1.

Since the voltage drop across the series resistor is constant for a given input voltage, the current

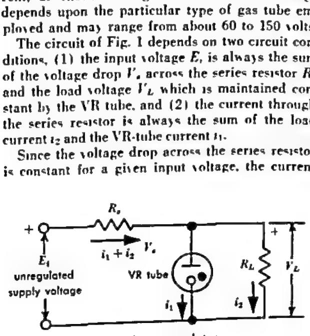


Fig. 1. Glow-tube voltage regulator.

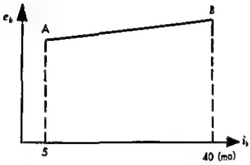


Fig. 2. A typical glow-tube volt-ampere characteristic. (From S. Seely, Electronic Engineering, McGraw-Hill, 1956)

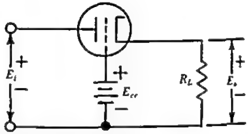


Fig. 3. Degenerative voltage regulator.

through the resistor must also be constant. Therefore, an increase in load current must be accompanied by an equal decrease in current through the VR tube and a decrease in load current must be accompanied by an equal increase in tube current.

If the input voltage varies, V_s must also vary since V_L is constant. The current through R_s must change if the voltage across it changes; therefore the total circuit current changes. Since there is no change in the load, the increase or decrease in current is entirely in the VR-tube current.

Vacuum-tube regulator circuits. One of the simplest of these circuits is the degenerative type using a series control tube as shown in Fig. 3. Since the cathode of the vacuum tube is at the potential E_o, the grid bias E_{cc} must be such as to make the grid potential slightly less than E_o. If the voltage E_o across the load should increase, the cathode of the tube is raised in potential. This tends to increase the resistance of the tube and thus increases the voltage across the tube. This will tend to bring the load voltage back to where it was initially. The circuit will, therefore, tend to compensate for load variations.

If the input voltage E_i increases, the plate potential of the tube increases, resulting in an increase of current in the tube. This raises E_o, making the cathode more positive. The increased grid-cathode potential of the tube tends to increase its resistance and its voltage drop. The result is that the increase in voltage drop largely overcomes the increase in the input voltage, and output voltage increases very little. The circuit will, therefore, tend to compensate for input voltage variations.

The regulator is not perfect; a small change in output voltage does occur when a much larger change in input voltage is pr[illegible] e ideal tube

would be one in which large changes in plate-to-cathode voltage result when a small change in the grid to cathode voltage occurs.

A more practical voltage regulator circuit is shown in Fig 4 The grid bias for the series control tube T is obtained from the gas-tube voltage regulator circuit of R_1, R_2, and VR similar to that of Fig. 1 The potentiometer R_2 allows the grid bias voltage of tube T to be adjusted so that the output voltage E_2 has the desired value The resis- ... the control grid of tube T ... sing

... es in ... The ... it of

... T will increase considerably the amount of ...

... 5, usually is a voltage regulator gas tube. If ... output voltage E_2 increases for some reason, ... voltage E_3 impressed on the grid of tube T_1 ... ll also rise, thus increasing the current flowing ... rough T_1 This increases the voltage drop across ... resistor R_3 and makes the grid of the series ... ntrol tube T_2 more negative. The voltage across ... then increases, tending to decrease the output ... oltage E_2 The net effect is to keep the output ... oltage nearly constant and thus compensate for ... oad variations In a similar fashion, if the unregu-

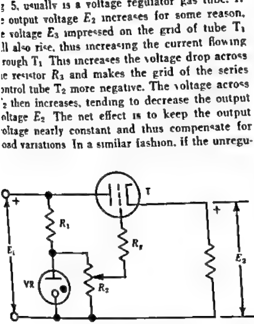


Fig 4 Series tube voltage regulator with grid bias controlled by a gas VR tube.

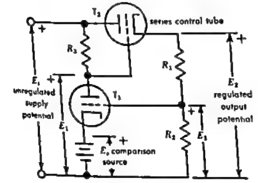


Fig 5. A basic electronic regulator circuit

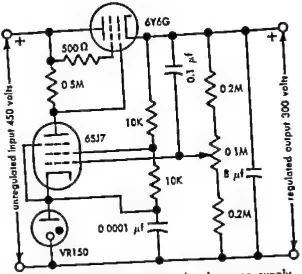


Fig 6 An electronically regulated power supply (From S. Seely, *Electronic Engineering*, McGraw-Hill, 1956)

lated supply potential E_1 increases, the voltage E_2 across the output would tend to increase. The sequence of events enumerated above would again occur, tending to increase the voltage across T_2 and again keep the output voltage practically constant. Thus the circuit will compensate for input voltage variations If the tube T_1 has sufficient gain, the voltage regulation will be nearly perfect In practice, T_1 usually would be a pentode tube, so that it would have sufficient gain The tube T_2 would be a power triode with low amplification factor and high current capacity or a beam power tube connected as a triode If the current taken by the load is too high for one of these triodes, two or more may be connected in parallel

An actual voltage regulator circuit using these principles is shown in Fig 6 The 0 1-microfarad (μf) capacitor connected from the cathode of the 6Y6G tube to the control grid of the 6SJ7 tube effectively connects the control grid of the 6SJ7 across the output voltage for rapid changes of output voltage This helps in counteracting voltage changes caused by pulses, square waves, high-frequency sine waves, and the like

A complete voltage-regulated power supply is shown in Fig 7 This shows a full wave rectifier, an *RC* smoothing filter, and a voltage regulator, or stabilizer. Additional gain is obtained in the voltage ...

... need only be an RC filter instead of a more expensive *LC* filter The 200-ohm resistor between the rectifier and the filter helps to limit the peak currents flowing in the tubes The internal resistance of the unit as viewed from the load is very low of the order of 1 or 2 ohms

A complete voltage-regulated power supply, employing a *p-n-p* power transistor 2N57 as the serie

tion of how sensitive its voltage will be to load changes under manual excitation control. A low value is desirable to give a constant voltage output. However, in large generators, a high value is often desirable, since it limits short-circuit currents and therefore protects the machine from damage.

[L.T.R.]

Voltage regulator

A device or circuit that maintains a voltage nearly constant over a range of variations of input voltage and load current. Voltage regulators are used wherever the unregulated voltage would vary more than that which can be tolerated by the electrical equipment using that voltage. Alternating-current distribution feeders use voltage regulators to keep the voltage supplied to the user within a prescribed range. Electronic equipment often has voltage regulators in the dc power supplies.

ELECTRONIC VOLTAGE REGULATORS

In much electronic equipment the dc power-supply voltage must remain constant in spite of input ac line voltage variations and output load variations. To provide a constant output voltage, two types of voltage regulator circuits are used in electronic power supplies, the gas-tube regulator circuit and the vacuum-tube regulator circuit. The voltage regulator circuit may also augment the filtering action of the power-supply filters, and it raises or lowers the internal impedance of the power supply as seen from the load. The price paid for the regulator action is a considerable drop in voltage, thus considerable power is lost in the regulator circuit. Most of the power is lost either in a regulator resistor or in the regulator tubes.

Gas-tube regulator circuit. One gas-tube regulator circuit is shown in Fig. 1. The tube used is a glow tube, which is characterized by almost a constant voltage over the specified range of tube current, as shown in Fig. 2. The constant voltage depends upon the particular type of gas tube employed and may range from about 60 to 150 volts.

The circuit of Fig 1 depends on two circuit conditions, (1) the input voltage E_i is always the sum of the voltage drop V_s across the series resistor R_s and the load voltage V_L which is maintained constant by the VR tube, and (2) the current through the series resistor is always the sum of the load current i_2 and the VR-tube current i_1.

Since the voltage drop across the series resistor is constant for a given input voltage, the current

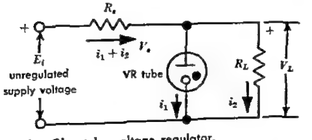


Fig. 1. Glow-tube voltage regulator.

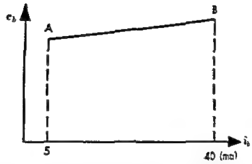


Fig. 2. A typical glow-tube volt-ampere characteristic (From S. Seely, *Electronic Engineering*, McGraw-Hill, 1956)

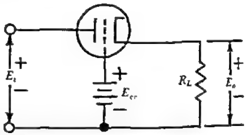


Fig 3 Degenerative voltage regulator.

through the resistor must also be constant. Therefore, an increase in load current must be accompanied by an equal decrease in current through the VR tube and a decrease in load current must be accompanied by an equal increase in tube current.

If the input voltage varies, V_s must also vary since V_L is constant. The current through R_s must change if the voltage across it changes; therefore the total circuit current changes. Since there is no change in the load, the increase or decrease in current is entirely in the VR-tube current.

Vacuum-tube regulator circuits. One of the simplest of these circuits is the degenerative type using a series control tube as shown in Fig 3. Since the cathode of the vacuum tube is at the potential E_o, the grid bias E_{cc} must be such as to make the grid potential slightly less than E_o. If the voltage E_o across the load should increase, the cathode of the tube is raised in potential. This tends to increase the resistance of the tube and thus increases the voltage across the tube. This will tend to bring the load voltage back to where it was initially. The circuit will, therefore, tend to compensate for load variations.

If the input voltage E_i increases, the plate potential of the tube increases, resulting in an increase of current in the tube. This raises E_o, making the cathode more positive. The increased grid-cathode potential of the tube tends to increase its resistance and its voltage drop. The result is that the increase in voltage drop largely overcomes the increase in the input voltage, and output voltage increases very little. The circuit will, therefore, tend to compensate for input voltage variations.

The regulator is not perfect; a small change in output voltage does occur when a much larger change in input voltage is pre[illegible] ideal tube

would be one in which large changes in plate-to-cathode voltage result when a small change in the grid-to-cathode voltage occurs

A more practical voltage regulator circuit is shown in Fig. 4. The grid bias for the series control tube T is obtained from the gas tube voltage regulator circuit of R_1, R_2, and VR similar to that of Fig. 1. The potentiometer R_2 allows the grid bias voltage of tube T to be adjusted so that the output voltage E_2 has the desired value. The resistor R_g is used to prevent the control grid of tube T from being over-driven positive and thus causing the output voltage E_2 to become too large.

The control exercised by most available tubes in the circuit of Fig. 4 is too small to be effective. The addition of a dc amplifier in the grid circuit of tube T will increase considerably the amount of control that can be exercised. A simplified circuit incorporating such an amplifier is shown in Fig. 5. The comparison source E_s, shown as a battery in Fig. 5, usually is a voltage regulator gas tube. If the output voltage E_2 increases for some reason, the voltage E_3 impressed on the grid of tube T_1 will also rise, thus increasing the current flowing through T_1. This increases the voltage drop across the resistor R_3 and makes the grid of the series control tube T_2 more negative. The voltage across T_2 then increases, tending to decrease the output voltage E_2. The net effect is to keep the output voltage nearly constant and thus compensate for load variations. In a similar fashion, if the unregulated supply potential E_1 increases, the voltage E_2 across the output would tend to increase. The sequence of events enumerated above would again occur, tending to increase the voltage across T_2 and again keep the output voltage practically constant. Thus the circuit will compensate for input voltage variations. If the tube T_1 has sufficient gain, the voltage regulation will be nearly perfect. In practice, T_1 usually would be a pentode tube, so that it would have sufficient gain. The tube T_2 would be a power triode with low amplification factor and high current capacity or a beam power tube connected as a triode. If the current taken by the load is too high for one of these triodes, two or more may be connected in parallel.

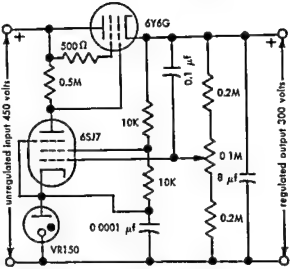


Fig. 6 An electronically regulated power supply. (From S. Seely, *Electronic Engineering*, McGraw-Hill, 1956)

An actual voltage regulator circuit using these principles is shown in Fig. 6. The 0.1-microfarad (μf) capacitor connected from the cathode of the 6Y6G tube to the control grid of the 6SJ7 tube effectively connects the control grid of the 6SJ7 across the output voltage for rapid changes of output voltage. This helps in counteracting voltage changes caused by pulses, square waves, and high-frequency sine waves, and the like.

A complete voltage-regulated power supply is shown in Fig. 7. This shows a full-wave rectifier, an *RC* smoothing filter, and a voltage regulator, or stabilizer. Additional gain is obtained in the voltage regulator by using a 6F8G containing two triode units in a cascode amplifier unit. Since the regulator circuit aids greatly in filtering, the filter used need only be an *RC* filter instead of a more expensive *LC* filter. The 200-ohm resistor between the rectifier and the filter helps to limit the peak currents flowing in the tubes. The internal resistance of the unit as viewed from the load is very low, of the order of 1 or 2 ohms.

A complete voltage-regulated power supply, employing a *p-n-p* power transistor 2N57 as the series

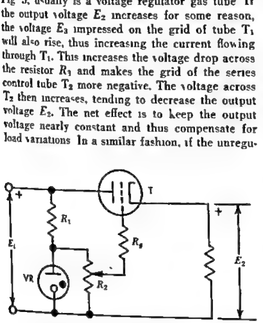


Fig. 4 Series tube voltage regulator with grid bias controlled by a gas VR tube

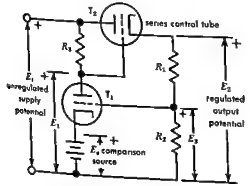


Fig. 5 A basic electronic regulator circuit

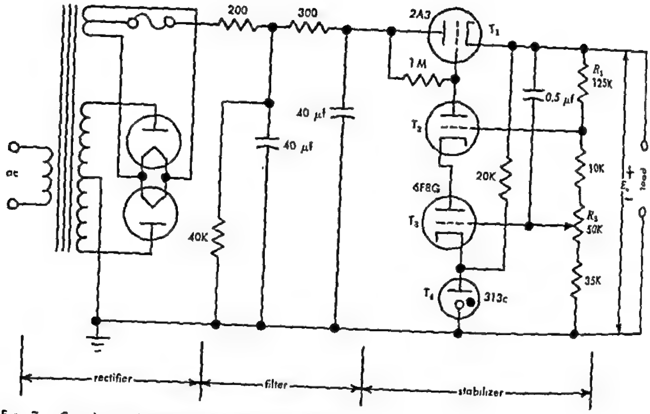


Fig. 7. Complete voltage-regulated power supply. (From H. E. Clifford and A. H. Wing, eds., Electronic Circuits and Tubes, McGraw-Hill, 1947)

voltage regulator, is shown in Fig 8. The two-stage direct-coupled amplifier uses an 1859 and a 951 transistor. The reference voltage of 15.6 volts is obtained from the reverse-biased breakdown characteristic of silicon-alloy junction diodes. See POWER SUPPLY, ELECTRONIC. [D.L.W.]

POWER-SYSTEM VOLTAGE REGULATORS

Voltage regulators are used on distribution feeders to maintain voltage constant, irrespective of changes in either load current or supply voltage. Voltage variations must be minimized for the efficient operation of industrial equipment and for the satisfactory functioning of domestic appliances, television in particular. Voltage is controlled at the system generators, but this alone is inadequate because each generator supplies many feeders of diverse impedance and load characteristics. Regulators are applied either in substations to control voltage on a bus or individual feeder or on the line to reregulate the outlying portions of the system. These regulators are variable autotransformers with the primary connected across the line. The secondary, in which an adjustable voltage is induced, is connected in series with the line to boost or buck the voltage.

A control and drive provide automatic operation. A voltage-regulating relay senses output voltage.

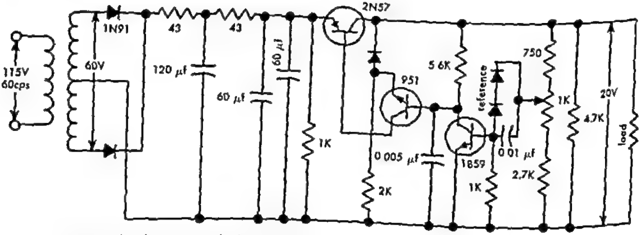


Fig. 8. Voltage-regulated power supply for transistor circuits. (From J. M. Carroll, Transistor Circuits and

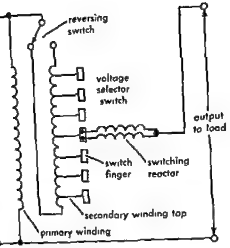


Fig. 9 Step voltage regulator circuit

When the voltage is either above or below the band of acceptable voltage maintained by the regulator, this relay causes the motor to operate and change the regulator position to raise or lower the voltage as required to bring it back within the control band It is desirable to maintain constant voltage at the average load center out on the feeder rather than at the regulator terminals Hence the control circuit includes a line-drop compensator with resistance and reactance elements that can be adjusted to represent line impedance. These impedances carry current proportional to circuit load, thereby simulating the voltage drop between the regulator and the load center, and modify the voltage sensed by the voltage regulating relay

Two principal types of feeder regulators are used: the step regulator, which provides increments or steps of voltage change, and the induction regulator, which provides continuous voltage adjustment.

Step voltage regulator. The transformation ratio of the autotransformer in this regulator is adjusted by a voltage selector switch which changes the secondary winding tap connected to the line Figure 9 shows the most commonly used circuit

Switching is performed without interrupting load current by means of the two switch fingers in the selector switch When the switch moves from the full tap position shown in Fig 9, one finger contacts the next tap before the other finger leaves the first tap; this constitutes a tap-bridging position Switching reactors limit the current circulating between the bridged taps. Most regulators are designed to operate continuously in these bridging positions (as well as the full tap positions) to provide voltages midway between the voltages of adjacent taps. Thus the voltage step between adjacent positions is half-tap voltage, and the number of winding taps required is half the number of operating positions in the boosting range The automatic reversing switch changes the polarity of the secondary relative to the primary, thus providing a bucking range equal to the boosting range.

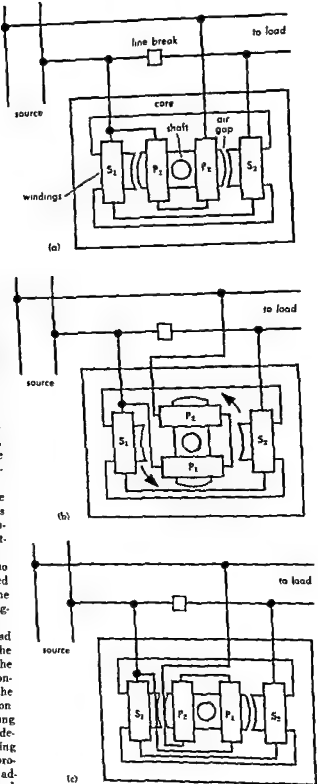


Fig. 10 Induction voltage regulator. (a) Full boost position. (b) Neutral position (c) Full buck position of rotor P coils are rotor primary coils, S coils are stationary secondary coils (From B G A Skrotski, ed , *Electric Transmission and Distribution*, McGraw-Hill, 1954)

Single- and three-phase designs are available with a range of $\pm 10\%$ in 16 or 32 steps. Other design variations may employ single-finger switching, and operate continuously only in full-tap positions.

Load tap-changing transformers are often used to provide both regulation and transformation from one voltage level to another. They are similar to step voltage regulators, except they have separate high and low-voltage windings.

Induction voltage regulator. This is similar in structure to a wound-rotor induction motor with the rotor restrained so that it moves only to adjust voltage. The primary winding on the rotor is magnetically coupled with the series secondary winding on the stator.

The principle of a two-pole single-phase regulator is illustrated in Fig. 10. Secondary voltage is continuously adjusted from full buck to full boost by changing the relative angular position of these windings through 180 electrical degrees.

Single-phase regulators require an additional permanently short-circuited rotor winding that is in space quadrature to the primary winding. Without this winding, the reactance of the regulator to the line current flowing in the secondary would be excessive in the neutral region between buck and boost. Three-phase induction regulators, if built on a three-phase core, do not require a short-circuited winding They inherently introduce phase shift between primary and secondary voltages and are no longer supplied for feeder regulation.

Other regulators. Other types of regulators are also used abroad One construction is a transformer structure with moving coils to change coupling, another has contacts moving over the exposed conductors to provide a large number of small, discrete steps.

Line voltage may be increased by drawing leading-power-factor current through the line reactance Static capacitors, shunt-connected in fixed or automatically switched banks, are often applied near the loads to raise voltage. The increase in voltage is not limited solely to the vicinity of the capacitor They also help compensate for the usual system condition of lagging power factor The application of series capacitors on fluctuating loads is increasing.

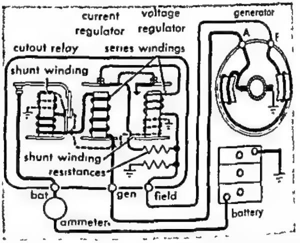
Infrequently, synchronous machines called synchronous condensers, over-excited to draw leading current, are connected to lines Their use has decreased with the availability of low-cost reliable static capacitors.

Conditions may exist where the inherent static capacity of the circuit is excessive. The line voltage may rise on cable systems or lightly loaded lines; shunt reactors may be used to neutralize the capacity of such systems. [D.D.M.]

Bibliography: J. M. Carroll, *Transistor Circuits and Applications*, 1957; H. E Clifford and A. H Wing (eds.), *Electronic Circuits and Tubes*, 1947, G. N. Patchett, *Automatic Voltage Regulators and Stabilizers*, 1954; S. Seely, *Electronic Engineering*, 1956.

Voltage regulator, automotive

A device in the automotive electrical system to prevent generator overvoltage. The device contains a set of spring-loaded contacts and a winding shunted across the generator. When the generated voltage exceeds the preset voltage the magnetic flux is sufficient to open the contacts, thereby inserting resistance in the generator field. Generator voltage is reduced, the contacts close, and the cycle is repeated. The contacts vibrate up to 200 times a second.



Wiring circuit of an automotive voltage regulator which includes a current regulator and a cutout relay (Delco-Remy Div., General Motors Corp.)

Included with the voltage regulator are a current regulator and a cutout relay on most applications (see illustration) The current regulator prevents generator overload The cutout relay opens the circuit when the generator is inoperative to prevent battery discharge through the generator [W.H.C.]

Voltage-multiplier circuit

A rectifier circuit capable of supplying a dc output voltage that is two or more times the peak value of the ac input Such circuits are especially useful for high-voltage, low-current supplies These supplies are usually lighter in weight, smaller in size and less expensive than the more usual half-wave and full-wave rectifier supplies. They require either no power transformer or a much smaller transformer,

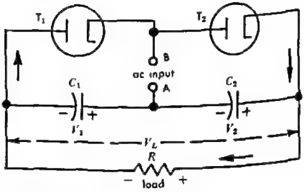


Fig 1. Full-wave voltage d

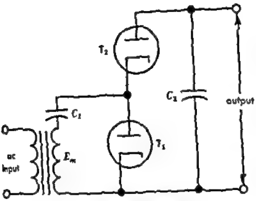


Fig 2 A half-wave voltage-doubling circuit

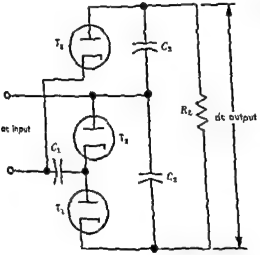


Fig 3 Voltage-tripler circuit, a voltage-doubler in series with a half wave rectifier

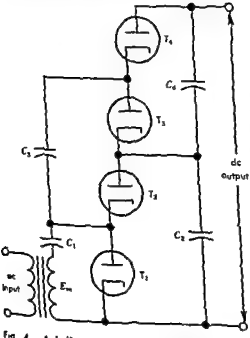


Fig 4 A half-wave potential-quadrupling circuit

but they have the disadvantage of requiring more rectifiers and capacitors. The most common circuits are the half-wave and full-wave voltage-doubling circuits, but tripling, quadrupling or even higher orders of multiplication are used

The full-wave voltage-doubling rectifier circuit is shown in Fig 1. When the ac input voltage is positive at terminal A, tube T_1 conducts, producing the voltage V_1 across capacitor C_1. On the other half cycle, tube T_2 conducts, producing the voltage V_2 across the other capacitor C_2. Both V_1 and V_2 will have a voltage approaching the maximum value of the ac input voltage, so the output voltage approaches twice the maximum value of the ac input voltage Since the output voltage receives one pulse every half cycle, the ripple voltage is similar to that of a single-phase full-wave rectifying circuit. *See* RECTIFIER.

The half-wave voltage-doubling rectifying circuit is shown in Fig 2. When the top terminal of the transformer secondary is positive, current flows through tube T_1 charging capacitor C_1 with its lower terminal positive in potential When the bottom terminal of the transformer secondary is positive, current flows through tube T_2 charging capacitor C_2 to nearly twice the maximum value of the secondary voltage. Only one pulse per cycle is received by the capacitor C_2, so the ripple voltage is similar to that of a single-phase half-wave rectifying circuit The full-wave doubler has the advantage of having a better voltage regulation and [illegible]

Fig 3, and a half-wave voltage-quadrupling rectifying circuit is shown in Fig. 4 Comparison of Figs. 2 and 4 show that a half-wave doubler can be transformed into a half-wave quadrupler by adding two capacitors and two tubes Theoretically, this may be continued indefinitely to achieve as high a multiplication as is desired. In practice, each stage added contributes slightly less to the output voltage than the preceding stage so a point is finally reached where an additional stage does not contribute enough to make it worthwhile *See* POWER SUPPLY, ELECTRONIC [D L W]

Voltage-regulator tube

A gaseous glow-discharge tube operating in the normal-glow region and therefore possessing an almost constant anode-to-cathode voltage drop as shown in Fig. 1 (*see* ELECTRICAL CONDUCTION IN GASES, GAS TUBE) The purpose of these tubes is to provide a substantially constant dc voltage from a fluctuating dc source

Typical construction of a voltage-regulator (*VR*) tube is shown in Fig 2*a*. The cathode is a large cylindrical electrode and the anode is a small wire, mounted concentric with the cathode The gap between the starting probe and the anode, combined with adjustment of the filling-gas pressure, permit the tube to be made with a starting, or breakdown, voltage only slightly higher than the glow voltage

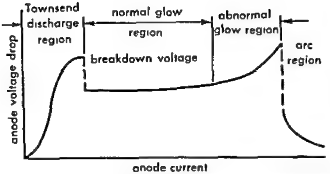


Fig 1. The gas discharge volt-ampere characteristics of a typical VR tube.

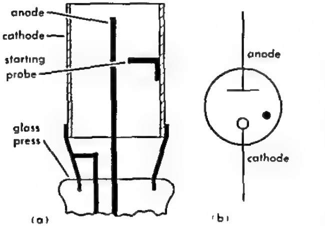


Fig 2. (a) Construction of a typical VR tube. (b) The common symbol for a VR tube.

The volt-ampere characteristics of a VR tube (Fig 1) show a slight increase in anode voltage drop for increasing current over the normal-glow region. For a typical tube, this voltage rise is about 5 volts. For use of this tube in voltage-regulator circuits *see* VOLTAGE REGULATOR

Voltage-regulator tubes are available in several different values of regulated voltage, such as 75, 90, 105, and 150 volts. These various voltages are attained by changing the nature of the cathode surface used and its treatment, and also the filling gas The regulating range is usually for currents of from 5-50 milliamperes (ma), but commercial tubes have been built for currents of 500 ma by using large cathodes. Voltage-reference tubes are small voltage-regulator tubes designed to maintain close voltage limits over a narrow range of current (4-5 ma). [J.D C.]

Bibliography: J. D. Ryder, *Engineering Electronics with Industrial Applications and Control*, 1957.

Volt-amperes

The product of rms voltage E and rms current I in an alternating-current circuit; also called the apparent power. This is a complex quantity; the real part is the active, or average, power in watts and the imaginary part is the reactive power in vars (volt-amperes, reactive). If voltage and current are in time phase, the apparent power in volt-amperes is equal to the active power in watts (Fig 1) For sine waves,

$$EI \cos \theta = \text{watts active (average) power} \quad (1)$$

and

$$EI \sin \theta = \text{volt-amperes reactive power} \quad (2)$$

where θ is the phase angle between the voltage E and the current I. *See* ALTERNATING-CURRENT CIRCUIT THEORY. [B L R]

Measurement of apparent power. The apparent power of a circuit is usually determined by measuring the effective voltage and current and obtaining their product. Since accurate instruments are available to measure both current and voltage, this method accurately determines the volt amperes of a circuit. *See* CURRENT MEASUREMENT; VOLTAGE MEASUREMENT.

Measurement of reactive power. The reactive power is of importance in analyzing completely the power requirements of a system Although this power does not produce work, it does produce conditions that make work possible. If, for example, a transformer is to operate or a motor is to rotate, this quadrature component, or reactive power, is required to magnetize the transformer core or the motor field.

Single-phase var measurement. From Eq (2) it can be inferred that the reactive power may be calculated in a single-phase circuit from observations of effective current and voltage and phase angle by conventional instruments. Alternatively, if active power is measured with a wattmeter, this observation with simultaneous observations of current and voltage can also lead to reactive power by the formula

$$\text{Vars} = \sqrt{(\text{volt-amperes})^2 - (\text{watts})^2}$$

Direct measurement of reactive power is possible using conventional wattmeters (*see* WATTMETER) by providing means for shifting the phase of the current in the potential circuit 90° with respect to the current in the load circuit. The means used may be a capacitance or an inductance With the electrodynamic wattmeter, usually a tapped inductance is substituted for the usual noninductive series resistance in the moving-coil circuit. This reactance, together with a shunted resistance, provides for adjustment to exact quadrature of the currents (Fig 2)

Similar phase-shifting means are applicable to the single-phase thermal wattmeter to accomplish var measurements. In either case this reactive circuit introduces an error at any frequency except

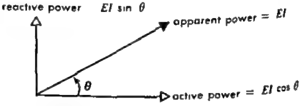


Fig. 1 Apparent, active, and reactive power relationships

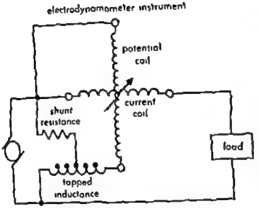


Fig 2 Single-phase reactive power measurement using an inductance

that for which it is adjusted Furthermore, the presence of harmonics also involves inaccuracy of measurement

Three phase var measurement In three-phase, three wire circuits, the measurement of reactive power is commonly accomplished by means of a phase shifting transformer In one form two autotransformers are connected in open delta, as shown in Fig 3 The wattmeter may be a two-element polyphase meter calibrated to read directly in vars The voltages applied to the potential coils of the meter are unchanged in magnitude but are shifted 90° from their positions when real power is being measured The current coils are usually supplied from current transformers The relations of currents and voltages may be shown to be as they would appear on separate instruments:

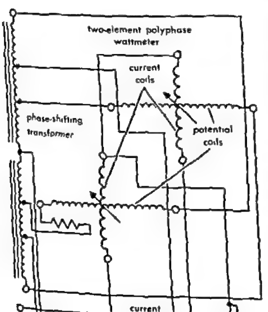


Fig 3 Three-phase reactive power measurement using phase shifting transformer

$$P_1 = EI \cos (60° - \theta) \qquad P_2 = EI \cos (120° - \theta)$$

and, thus,

$$P_1 + P_2 = \sqrt{3}\, EI \sin \theta$$

It is assumed that phase-to-phase voltages are symmetrical in this method; therefore, any lack of symmetry results in errors in readings.

Var measurements and volt-ampere measurements are generally subject to the same classes of error as measurements of real power with wattmeters of the particular type used. *See* ELECTRIC POWER MEASUREMENT. [C.A.M.]

Bibliography: F. K. Harris, *Electrical Measurements*, 1952. F. A. Laws, *Electrical Measurements*, 2d ed., 1938.

Voltmeter

An instrument for the measurement of potential difference between two points, in volts Derivatives of the voltmeter are the microvoltmeter, millivoltmeter, and kilovoltmeter, for measurement of voltages with a measurement span of 1,000,000,000:1. Voltmeters are connected between points of a circuit, between which the potential difference is to be measured (*see* VOLTAGE MEASUREMENT). The accuracy rating is usually stated in terms of the full-scale reading

Classifications. Voltmeters are classified in respect to the kind of voltage to be measured and the kind of mechanism used, as follows. (1) for measurement of direct voltage, (*a*) the moving-coil permanent-magnet voltmeter, (*b*) the fixed-coil moving-magnet voltmeter, and (*c*) the electrostatic voltmeter, (2) for measurement of alternating voltage, (*a*) the *fixed-coil moving magnet voltmeter*, (*b*) the fixed-coil moving-iron voltmeter, (*c*) the thermovoltmeter consisting of a thermal converter and dc millivoltmeter, and (*d*) rectifier and vacuum-tube voltmeters

Voltmeters, in general, are secondary instruments consisting of voltage-to-current transducers and mechanisms which respond to milliamperes or microamperes In the simpler forms, the transducer consists of a resistor of constant value, which may have taps for several voltage ranges *Since by* Ohm's law the current is proportional to the potential difference, calibration of the combination using a primary standard, such as the potentiometer and standard cell, is valid. In the more specialized forms, the transducer consists of a thermal converter (ac to dc), rectifier, or amplifier The electrostatic instrument is the only voltmeter which measures potential difference directly.

Moving-coil voltmeter. This voltmeter consists of a permanent magnet, moving coils pivoted in jewel bearings, control springs, and pointer, as shown in Fig. 1. The current I is passed through the moving coil of N turns via the two control springs which apply a restraining torque T, equal to $K_s\theta$, where K_s is the spring constant and θ is the angle of

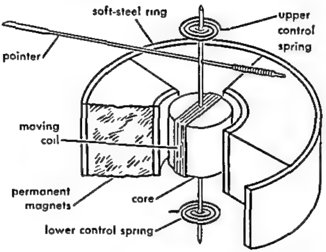


Fig 1. DC permanent-magnet, moving-coil voltmeter. (*General Electric Co*)

deflection. The deflecting torque T_r is equal to the product of the magnetic moment of the coil and the field intensity B, thus

$$T_r = NIwlB \quad (1)$$

where w and l are coil width and length, respectively. The deflection θ resulting from the current I is

$$\theta = BNIwl/K_s \quad \text{radians} \quad (2)$$

Since the magnetic field is usually uniform, the voltmeter scale is usually calibrated in uniform intervals.

The moving system must be damped to eliminate unwanted oscillations of the pointer. This is accomplished electromagnetically by a damping shell, usually aluminum, on which the moving coil is wound. As the coil moves, the shell moves in the magnetic field. A voltage and current are induced in the shell in the direction of the winding. This current reacts with the magnetic field to produce a braking action proportional to the angular velocity of the moving system. The amount of damping is controlled by the cross section and conductivity of the shell. It is customary to damp voltmeters so that a small but definite overshoot will result if full voltage is suddenly applied. In some instances, the shell is replaced by an equivalent coil and in others the instrument may be damped by the action of the driving coil itself. The latter condition is achievable only when the loop circuit resistance is low.

Permanent-magnet moving-coil instruments are made in several forms, of which four common ones are illustrated in Fig. 2.

The first three figures illustrate the evolution of the magnet structure, beginning with the expanded U-shaped magnet constructed of chrome or tungsten steels, and extending to the internal cylindrical magnet made with Alnico material of high energy content. These mechanisms have scale lengths of about 90°. The fourth shows an annular magnet and offset coil with a scale length of 240°. Voltmeters and their derivatives are available in ranges of 10 mv–750 volts, with resistances of 100–10,000 ohms/volt. Maximum errors range from 0.1% of full scale for secondary standards to 2.0% of full scale for panel-type instruments.

Moving-magnet voltmeter. This voltmeter, illustrated in Fig. 3, has the magnet attached to the pivoted shaft which carries the pointer. The magnet is surrounded by the field coil and aligned by a control magnet which supplies the restoring torque. A fixed copper frame damps the movement electromagnetically, and the magnetic shield minimizes the effect of external fields, as well as providing a path for the internal fluxes. As current is applied to the field coil, the coil and restoring-magnet fields produce a resultant field. The rotor aligns itself with this resultant field and indicates the magnitude of the current and thus the potential difference. The construction shown is used in commercial 2½-in. and 3½-in. diameter panel instruments of 2% accuracy rating (maximum error of 2%).

Electrostatic voltmeter. The electrostatic voltmeter action is based on the force of attraction or repulsion between two charged conductors, such as the plates of a variable air capacitor. The moving plate, when charged, tends to move so as to increase the capacitance between it and the fixed plate. If this capacitance is C farads, the voltage is E volts and the spacing is s meters, the energy W in joules of the capacitor is expressed as

$$W = \tfrac{1}{2}CE^2 \quad (3)$$

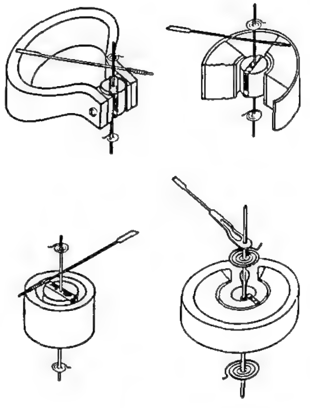


Fig 2. Four forms of dc moving-coil mechanisms. (From *I F Kinnard, Applied Electrical Measurements*, Wiley, 1956)

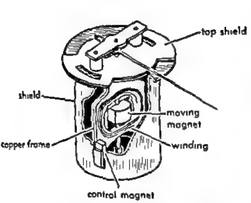


Fig 3 Moving-magnet dc voltmeter. (General Electric Co)

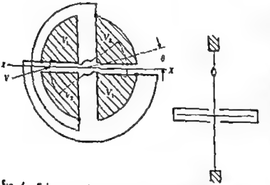


Fig 4 Schematic drawing of the quadrant electrometer (From I F Kinnard, Applied Electrical Measurements, Wiley, 1956)

If the upper plate is moved vertically a distance ds while the voltage is held constant, an energy change dW takes place, numerically equal to the work done in moving the plate. The resultant force F is

$$F = dW/ds = (E^2/2)\, dC/ds \qquad (4)$$

Thus, the force acting on the upper plate is proportional to E^2 times the space rate of change of capacitance For a rotatable system the corresponding torque is

$$T = (E^2/2)\, dC/d\theta \quad \text{radians} \qquad (5)$$

The quadrant electrometer. Fig. 4, is a useful embodiment of this principle The four quadrants compose the fixed capacitor plates and surround the movable vane suspended by a conducting torsion fiber at the center of the system Opposite quadrants are *electrically* connected and potentials V_1 and V_2 are applied In the heterostatic method, the vane is *independently* energized to potential V_3. The capacitance effect causes the vane to turn out of one *pair of* quadrants and into the other This movement actuates the indicating needle and the *deflection* of the needle is directly proportional to the voltage *difference* $(V_1 - V_2)$ between the two sets of quadrants.

In the idiostatic method, the vane is connected to quadrant 1. This method has the advantage of dispensing with the auxiliary voltage, and the deflection is proportional to the square of the voltage difference.

The Lindemann electrometer, Fig 5, is a variant of the quadrant electrometer, designed for portability and insensitivity to changes in position. The quadrants 1 and 2 are two sets of plates about 6 mm apart and mounted on insulating quartz pillars. A taut silvered quartz suspension fixes the center of rotation of the moving system. When voltage is applied to the needle, the needle rotates toward the oppositely charged plates. This movement is observed through a microscope. This electrometer has a low capacitance of about 1 micromicrofarad ($\mu\mu$f), and currents of the order of 10^{-15} amp can be observed with it

Electrodynamic ac voltmeter. This voltmeter has two coils connected in series through a resistor to the source, the potential of which is to be measured, as shown in Fig. 6. The fixed coil provides the field in which the moving coil, supported by a pivoted shaft, operates The deflection of the moving system is restrained by a control spring. When the circuit is closed, the current which flows through the coils produces a deflecting torque which moves the indicating needle. This current I produces a deflecting torque T_e

$$T_e = K_1 I^2\, dM/d\theta \qquad (6)$$

where M is that portion of the total field flux which links the moving system, K_1 is a constant of proportionality, and θ is the angle between moving

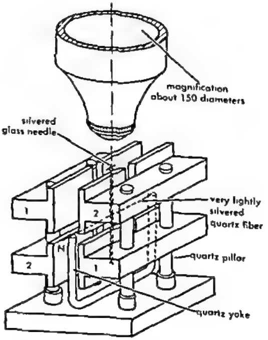


Fig 5. Lindemann electrometer—case omitted (From F A. Laws, Electrical Measurements, 2d ed., McGraw-Hill, 1938)

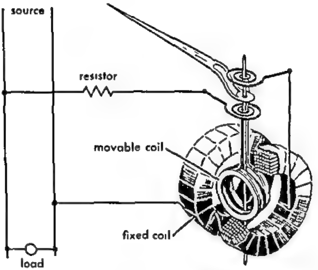


Fig 6. Electrodynamic mechanism used as voltmeter. (*General Electric Co.*)

coil and fixed coil Since the control spring produces a restoring torque T_r at deflection α,

$$T_r = K_S\alpha \qquad (7)$$

where K_S is the constant of the spring. The moving system will come to rest at an angle where deflecting and restoring torques are equal, and the indication is

$$\alpha = \left(\frac{K_1}{K_S}\right)\left(\frac{I^2\, dM}{d\theta}\right) \qquad (8)$$

The scale distributions of electrodynamic voltmeters are in general contracted at low voltages and expanded at higher voltages This is due in part to

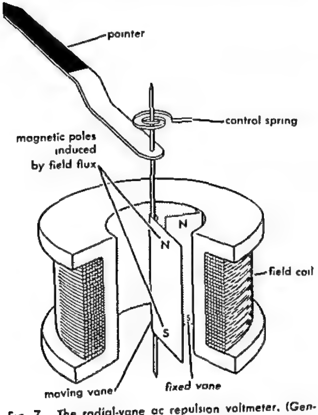


Fig 7. The radial-vane ac repulsion voltmeter. (*General Electric Co*)

the necessary square-law response and in part to designing for scale legibility over the upper half of the scale. The inherent accuracy of the electrodynamic voltmeter is excellent, and this construction is used in secondary standards of 0 1% error rating as well as portable voltmeters of 0 25 or 0.50% error rating Voltage ranges of 10-750 volts are available in commercial instruments Refinements in design include magnetic shielding, damping of the moving system, and compensation for temperature and frequency errors.

Moving-iron voltmeter. There are many forms of this voltmeter. Figure 7 is representative Here, the field coil is connected to the line through a series resistor (not shown). Inside the field coil are two

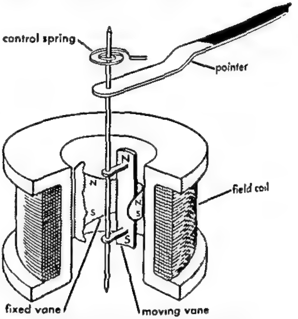


Fig 8. The concentric-vane ac repulsion voltmeter (*General Electric Co.*)

rectangular vanes, one fixed and the other attached to the shaft which carries the pointer. When a current traverses the coil the vanes are similarly magnetized, and the moving vane is repelled from the fixed vane, deflecting the pointer in a clockwise direction The deflecting torque T_e is

$$T_e = K\frac{I^2\, dL}{d\theta} \qquad (9)$$

where I is the current traversing the coil, L is the inductance of the field coil, and θ the angle between the vanes. Since the coil inductance is a maximum when the vanes are at maximum separation, the moving vane will, unless restrained, move 180° away from the fixed vane The control spring exerts a restraining torque

$$T_r = K_S\alpha \qquad (10)$$

or the deflection is

$$\alpha = \frac{K}{K_S}\frac{I^2\, dL}{d\theta} \qquad (11)$$

The moving-iron instrume — he electrody-

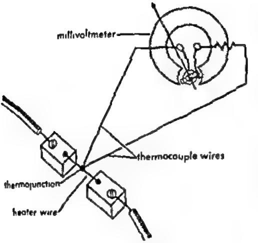


Fig 9 Elementary thermovoltmeter circuit. (*General Electric Co*)

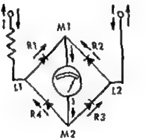


Fig 10 Operating principle of the bridge-type rectifier voltmeter Solid arrows show direction of current during positive half-cycle of ac input, dashed arrows show direction of current during negative half-cycle (General Electric Co)

namic voltmeter, when properly calibrated indicates true rms volts The scale length and distribution can be controlled over wide limits by suitable configuration of the fixed iron. Figure 7 illustrates an arrangement for scale lengths of 90–100°. Figure 8 illustrates a circumferential fixed vane design used in long-scale (250°) switchboard voltmeters Moving-iron voltmeters with magnetic damping are available in commercial 2½-in. and 3½-in panel instruments (2% error), 4-in. and 6-in switchboard instruments (1% error), and portable instruments of ¼ or ½% error. Scale ranges vary from 0–10 to 0–750 volts. Instruments of 1% accuracy and higher are magnetically shielded. Power [illegible]

[illegible] junction and dc millivoltmeter (Fig 9). A current from the voltage source is passed through a resistor (not shown) and a fine vacuum-enclosed platinum heater wire. A thermocouple, attached to the midpoint of the heater, generates millivolts proportional to the temperature *See* THERMOCOUPLE.

The thermovoltmeter is a true rms instrument, in which the squaring, integration, and averaging are accomplished thermally rather than electromagnetically The inductance of the heater wire is negligible, and the thermovoltmeter is especially suitable for measurement at high frequencies Its sensitivity, while greater than that of electromagnetic instruments, is limited by the current necessary to heat the wire, usually several milliamps.

Rectifier voltmeter. This voltmeter consists of a dc milliammeter calibrated in volts and connected to the voltage source through a rectifier bridge (Fig 10) The individual rectifiers are connected so that the current through the milliammeter is always in the same direction This meter measures

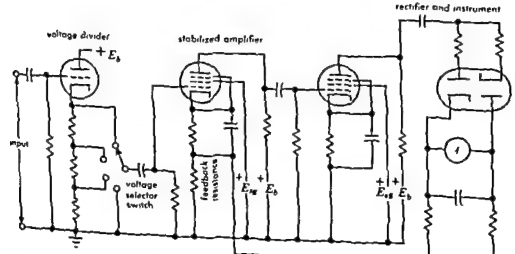


Fig 11 Circuit for amplifier-rectifier electronic voltmeter (From I F. Kinnard, *Applied Electrical Measurements*, Wiley, 1956)

the average value of the current. While not an rms instrument, the rectifier voltmeter may be calibrated in rms volts for sine waves and will measure sine-wave voltages accurately. Harmonics in the voltage wave cause appreciable errors, but this circuit has the advantage of low power consumption. Commercial forms of rectifier voltmeters having resistance of 1000 ohms/volt and ratings of 10–750 volts are available.

Electronic ac voltmeter. This voltmeter consists of a dc milliammeter calibrated in volts and connected to an amplifier-rectifier circuit, of which Fig. 11 is typical. The input voltage is connected to the grid of the first triode. A voltage-range selector switch is connected to the tapped cathode resistor. The two stages of amplification are resistance-coupled to cover a wide frequency range. The amplifier output is applied to a rectifier section consisting of two diodes, milliammeter, and feedback resistor. *See* DIGITAL VOLTMETER; VACUUM-TUBE VOLTMETER.

By use of adequate feedback, errors due to changes in power-supply voltage and tube characteristics are minimized. Electronic voltmeters usually respond to average voltage but are calibrated in rms values for sine waves. The addition of the amplifier greatly increases the sensitivity and range of measurement, 0.001–300 volts at frequencies of 20 cps–2 Mc, and with input impedances of several megohms [A.J.C.]

Bibliography: I. F. Kinnard, *Applied Electrical Measurements*, 1956; F. A. Laws, *Electrical Measurements*, 2d ed., 1938; Meter and Instrument Dept., General Electric Co., *Manual of Electric Instruments*, 1952.

Volt-ohm-milliammeter

A self-contained test instrument for measuring a wide range of voltages (both ac and dc), resistances, and currents (usually dc only), particularly in radio, TV, and electronic servicing

All readings are obtained on a single multiscale indicating instrument The selection of the correct ranges and function is accomplished either by means of one or two rotary switches, by multiple pin jacks, or by a combination of both.

Instrument sensitivity for dc-voltage measurements is usually 20,000 ohms per volt, although the figure ranges from 1000 to 100,000, for ac-voltage measurements the sensitivity ranges from 1000 to 20,000 ohms per volt. Accuracy is usually about 3% for dc, and 5% for ac measurements

The terms multimeter and analyzer (or circuit analyzer) are quite commonly used as synonyms for volt-ohm-milliammeter *See* AMMETER; OHMMETER; VOLTMETER. [I F K]

Volume unit (VU)

A unit f [illegible] the magnitude of a complex electric [illegible] or music. [illegible] times the common logarithm of the power ratio p_2/p_1, where the reference power level p_1 is selected as 1 mw (0.001 watt) and p_2 is the signal power level. Meters which give readings in volume units, called VU meters, are widely used for monitoring radio broadcasts and for sound recording. *See* BEL.

Volumetric analysis

Volumetric analysis, or possibly more correctly titrimetric analysis, is one of the major divisions of quantitative analytical methods In this method as the names imply, a titration is made and the amount of the desired component is determined indirectly from the volume and concentration of solution used in the titration.

In most volumetric procedures, the sample to be analyzed is weighed, or an accurately measured volume of a solution of the sample is taken The sample is dissolved in an appropriate solvent and subjected to some preliminary treatment or separation if necessary, and then the titration is performed on the resulting solution.

During the titration, a chemical reaction occurs between a constituent of the titrant solution and the substance to be determined. In order to be suitable for volumetric titrations, these reactions must be rapid and quantitative. That is, it is necessary to know exactly how many molecules or ions of the substance to be determined react with each molecule or ion of the reagent in the titrant In volumetric analysis, it is also necessary to determine when the number of reagent ions or molecules equivalent to the substance being determined has been added. This point in the titration is called the equivalence point The determination of the amount of titrant corresponding to the equivalence point is usually done either by visual means with colored indicators or reagents, or by some physicochemical measurement. When an indicator changes color, the end point of the titration is reached, and at that point, the volume of titrant solution added is measured Not always, however, are the end point and equivalence point of the titration identical An indicator, for example, may consume a small amount of titrant before it will change color, or the indicator may not change color until a concentration of reagent considerably higher than that corresponding to the equivalence point is present in solution In either of these cases, the point where the indicator changes color will not be the same as the true equivalence point For this reason, it is very essential that the choice of indicator or the method of detecting the end point in a titration be considered carefully in order to minimize the discrepancy between the detected end point and the true equivalence point

In order to calculate the amount of substance equivalent to the volume of titrant used in the titration, it is necessary to know accurately the concentration of reagent in the titrant This concentration is usually determined by weighing the amount of reagent added to a definite volume of titrant solution, by titrating a weight of sample of known composition with the titrant, or by titrating a known volume of another solutio[illegible] composition

From the volume of titrant used to reach the end point and the concentration of reagent in the titrant the weight of substance to be determined in the sample is then calculated

In volumetric analysis, the volume of titrant is usually measured although for the most accurate work the weight of titrant is measured Both of these procedures fall into the classification of titrimetric analysis *See* BURET, PIPET; QUANTITATIVE CHEMICAL ANALYSIS, SOLUBILIZING OF SAMPLES; STOICHIOMETRY, TITRATION, VOLUMETRIC FLASK. [C E.B.]

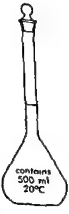
Volumetric efficiency

In describing an engine or gas compressor, volumetric efficiency is the ratio of volume of working substance actually admitted, measured at specified temperature and pressure to the full displacement volume For a liquid fuel engine such as a diesel engine volumetric efficiency is the ratio of volume of air drawn into a cylinder to the cylinder displacement *For a gas-fuel* engine such as a gasoline engine with carburetor, volumetric efficiency is based on the charge of fuel and air drawn into the cylinder

Volumetric efficiency of naturally aspirated automobile and aircraft reciprocating engines may be 85–90% at rated speed Supercharging increases volumetric efficiency, giving values *over 100%* At low speeds volumetric efficiency is high, at high speeds resistance to air flow in the manifolds and valve ports decreases volumetric *efficiency* Air compressors, refrigerator compressors, and dry vacuum pumps are generally specified for a volumetric efficiency of 60–90% dependent upon the compressor ratio, type of valve valve gear, and machine speed [N M]

Volumetric flask

A long narrow-necked glass flask which has been calibrated to contain a given volume of liquid when filled so that the bottom of the meniscus is tangent to the ring which has *been etched* on the neck of the flask Some volumetric flasks have a second calibration mark higher on the neck of the flask, which is used when the flask is needed to deliver a given volume of liquid. The marks etched on volumetric flasks are accurate only at the temperature at which these calibrations were made, because the volume of a flask increases slightly with an increase of temperature and vice versa. Volumetric flasks are used when it is necessary to dilute a weight or a volume of a substance to some accurately known volume *See* CONCENTRATION SCALES; TITRATION; VOLUMETRIC ANALYSIS. [C.E B.]



lumetric flask

Vortac

A ground radio station consisting of a co-located vhf omnidirectional radio range (VOR) and Tacan facility. This station permits obtaining polar coordinates by the use of a VOR receiver and distance measuring equipment (DMET) or by Tacan equipment. *See* RADIO RANGE; TACAN. [P C.S]

Vortex

A line vortex in two-dimensional fluid flow produces a flow or circulation around the line

Free vortex. Consider the effect of rotating a right-circular cylinder of radius r_0 about its axis with a peripheral velocity v_0 in a fluid otherwise at rest The fluid in contact with the surface of the cylinder rotates with the cylinder Fluid at greater radius is also set in motion in concentric circles with velocity diminishing as the radius increases This type of fluid motion in which the velocity varies inversely as the radius, is called a free vortex If the cylinder is reduced to zero radius in such a manner that $v_0 r_0$ remains constant in the limit as r_0 approaches zero, a line vortex results. The velocity at the line is infinite, so the line itself must be considered as a singular line, to be excluded from the actual fluid

Examples of vortices occur frequently in nature The tornado is an example of a free vortex, with high velocities near its center, and correspondingly low pressure intensities The waterspout is its counterpart over water.

The fluid motion in the case of a line vortex in an ideal (frictionless and incompressible) fluid is irrotational; that is, its motion may be *described* in terms of a velocity potential

Vortex tube. If a small spherical particle of a frictionless fluid could be *considered* as suddenly solidified its resulting rotation could be expressed by a vector parallel to the axis of rotation with its length proportional to the *angular velocity*, and with its direction indicating the sense of rotation by the right-hand rule When the rotation vector is everywhere zero *throughout a* region of fluid, the motion of that fluid is irrotational When some finite fluid regions have nonzero values of the rotation vector, then this fluid has *vorticity* A vortex line is a line drawn through the fluid such that it is everywhere tangent to the rotation vector A collection of vortex lines through a small closed curve defines a vortex tube, which has certain special properties

1 The circulation about a vortex tube is every-

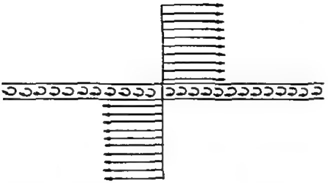


Fig. 1. Vortex sheet formed between oppositely directed streams.

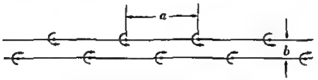


Fig. 2. Kármán vortex street consists of alternate vortices of opposite rotational sense

where the same along its length Circulation is defined as the line integral of the velocity vector around a closed path.

2. The vortex tube cannot end in the fluid It must either extend to a boundary or close upon itself.

3. Vortex lines move with the fluid. Vorticity of a fluid is a property of the fluid itself and not the space it occupies

A smoke ring is a practical example of a closed vortex tube. *A circular vortex tube in otherwise* still fluid will translate perpendicular to the plane of the ring without change in size.

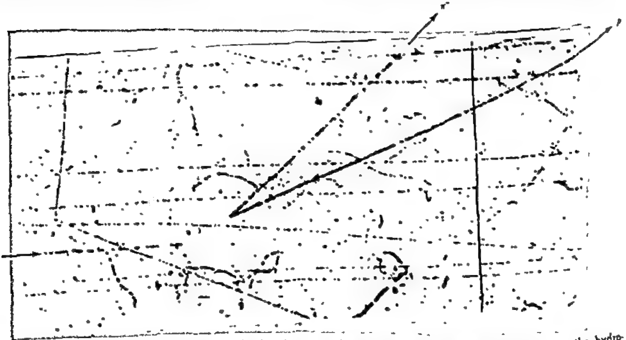
Vortex arrays. A discontinuity in fluid velocity along a surface, such as slippage of one layer of fluid over another, may be handled as a vortex sheet in an otherwise continuous flow. In this case all *vortex lines are in the surface (Fig 1). A practical* case of a vortex sheet is the flow downstream from an airfoil when the velocity leaving the upper surface is higher than the velocity leaving the under surface.

When real fluid flows around a body, such as wind blowing across a cable, the fluid rotation in the boundary layer causes vortices to form along the downstream side of the body. For certain conditions they remain near the body and are referred to as bound vortices For higher velocities the vortices form, grow and are then systematically shed from the downstream side of the body, forming a vortex street (Fig. 2). The unsymmetrical case shown is called the Kármán vortex street after Th von Kármán who first identified them and showed that the motion is stable when $b/a = 0.281$ *See* FLUID-FLOW PRINCIPLES. [V.L.S.]

V-particle

The name first used for the unstable particles whose decay is responsible for the production of characteristic V-shaped tracks observed in cloud chambers exposed to cosmic radiation

V-particle events were first reported in 1947 by the cosmic ray group of Manchester University and are now known to represent the decay of *K*-mesons or hyperons (*see* HYPERON; MESON). Most striking are the neutral V-events A neutral unstable particle often travels several centimeters



The V-particle event indicated in the bubble-chamber photograph above represents the decay of a Λ hyperon and points back to the point of disappearance of a π^- meson with kinetic energy of 1300 Mev entering from the left. At this point the π^- meson was absorbed on interacting with a proton (*p*) in the hydrogen of the chamber This particular event was actually established to be the first clear example of the reaction $\pi^- + p \rightarrow \Lambda + K_2^0$. (Dr N P. Samios)

before decaying, the two charged particles resulting from this decay are projected forward and produce tracks forming a V whose head points back to the nuclear collision in which the neutral V-particle was formed (see illustration). A charged V-event, caused by the decay of a charged unstable particle giving one charged secondary, appears as a sudden change of direction in the track of the incident particle, at which point some change in the appearance of the track (for example, thickness, structure, or curvature) generally occurs. See COSMIC RAYS. [R H D]

Vulture

Any of several carrion-feeding birds related to the hawk. Old World vultures belong to the family Accipitridae, but New World vultures belong to a separate family, the Cathartidae, characterized by unfeathered heads. There are six species of the Cathartidae, three in the United States. Vultures are not nearly as common as formerly. Largest and rarest is the California condor, *Gymnogyps californianus*, somewhat like the turkey vulture, but with an orange or yellow head and a wingspread of 8½–10½ ft. Most common is the turkey vulture, *Cathartes aura*, with a red head and a relatively long, narrow, somewhat rounded tail. *Coragyps atratus*, the black vulture, has a short, broad, square tail and a black head. Black vultures tend to concentrate along the coast but are found over the southern half of the United States. See FALCONIFORMES. [J D B.]

The turkey vulture, *Cathartes aura*, length to 32 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

W

Wage incentives to Wurtzite

Wage incentives

Plans that pay more money for more than stated output. Older wage incentives, like straight sales commission and straight piecework, had no stated output. Today, with minimum wage laws, the added earnings may begin at the minimum wage. Modern wage incentives usually have guaranteed wages comparable with community pay levels; incentive payments are above these amounts and are generally at the same rate. The most common plan is called *one-for-one*, meaning 1% more pay above *base rate* for 1% more than standard production.

Suppose the base rate is $2.40 per hour or 4¢ per minute. Assume that standard time is 1 *min* per piece. Then with a piece work plan the price is 4¢ per piece and 60 pieces per hour is standard output. Extra pay is 4¢ each for all acceptable quality production over 60 pieces per hour.

More plans use standard time per piece than money per piece because wage rates change so often. Besides, production planning and overhead costing are more easily and correctly carried on with time measures.

Thus the better plans are explained as $2.40 per hour worked plus 4¢ for each *extra minute* of production above 60 standard minutes *per hour*. Standard times may be expressed *in standard hours*. Extra production is usually computed as *total standard time output* for the day minus total actual time on measured work, credit being recorded for any delays when several jobs are done *in one day* (see table).

Simplified example of one-for-one incentive calculation

Time-card record	Output	Delays
80 pieces × 1.4 min	112	
Wait for foreman		4
20 pieces × 5.0 min	100	
Wait for inspector		3
100 pieces × 2.0 min	200	
Wait for material		9
50 pieces × 3.2 min	160	
Totals	572	16
Base pay = 8 hr × $2.40 = $19.20	464*	
Incentive premium = 108 × $.04 = $4.32	108	

* 480 minutes in 8 hour day less 16 minutes delay

Standard time and piece work are the usual wage-incentive plans, with standard time increasingly predominant. The two usual plans, however, do exist with variations of one-for-one. A few pay less, by design, where some part of the incentive is reserved for payments to indirect service people and supervisors. A few plans pay less than one-for-one because piece rates or premium rates per *minute* have not been increased proportionally with wage rates (Fig. 1). The opposite is true of those few plans that use a *lower* than going wage rate with a higher incentive rate. These may be called variations of the Gantt step bonus plan (Fig. 2).

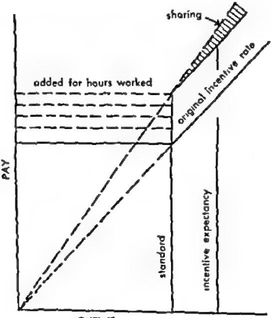


Fig. 1

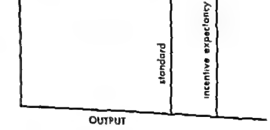


Fig. 2

Many incentive plans are based on "guesstimates." Increasingly, however, these are being replaced by standards based on work measurement. Newly installed plans are almost always based on work measurement. Similar plans are being installed to pay for extra output by indirect people in maintenance, toolroom, janitor, service and office work. *See* WORK MEASUREMENT.

Wage incentives extend rewards for increased *output to productive personnel just as salesmen* have long been rewarded for increased sales by commission payments. For salesmen, sales quotas are a measure of standard performance and sales are measured in dollars. Some incentive plans ex*tend throughout a company by being based on profit* as a measure of achievement. Personnel may participate in profit sharing plans in proportion to their base pay and longevity with the company. For executives, profit is an appropriate basis for a bonus because it reflects over-all performance. Similarly, many companies reward employees who improve plant performance in ways beyond their direct personal effort by paying a bonus for suggestions that improve methods, equipment, or product; increase rates of output, or reduce costs of materials, supplies, or services.

Except for salesmen, wage incentives have been utilized more with factory people than elsewhere. The prime reason is that during most of our industrial growth, production capacity and sometimes skilled manpower have lagged behind consumer increases. The increased production to meet demands of World Wars I and II provided our most striking examples of greatly accelerated installations of wage incentives.

Average increase in production above past performance reported for 234 installations in World War II was 41.5%; the range was 0 to 103%. The most often quoted past performance is 60% of standard. Usual expectancy of incentive earnings is 25% above standard These figures suggest that output can be doubled (125/60) Such increases have marked effects on wage costs. They spread overhead expenses and bring down direct labor costs *in relation to the difference between past per*formance and standard output

Better maintenance, as with equipment, can improve and extend the effectiveness of incentives In contrast with equipment, however, incentive plans are not *impersonal. They relate to people* Emotions, pride, skill and prejudice are mixed with the question of what is a fair day's work. Increases in output, demanded by progress and improvement, are more noticeable with incentives because standards or [illegible] advanta [illegible] Wage in[illegible] plied skill. This is in accord with our competitive system and our recognition of the individual. (*See* METHODS ENGINEERING.) [P.C.]

Bibliography: P. Carroll, *Better Wage Incentives*, 1957; P. Carroll, *How to Control Production Costs*, 1953.

Wake (seagoing vessels)

The disturbance astern created by a ship in motion. The disturbance is generated in water by the passage of the underwater body of the ship and its propulsion devices, and in air by the above-water body and superstructure. Wake is the composite phenomenon resulting from four principal contributors: (1) boundary layer flow created by the fric*tional resistance of the hull*, (2) potential flow caused by the displacement of the streamlines by the body, (3) standing surface wave patterns caused by gravity effects, and (4) slip streams from propellers.

Frictional wake. *This originates from the shear*ing resistance of a viscous fluid to the passage of a body, and appears as the growth of a boundary

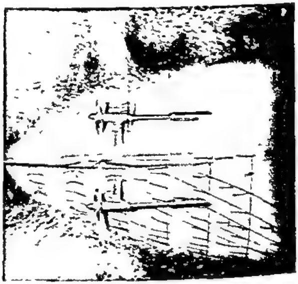


Fig. 1. Wake at the stern of a supertanker model Nylon tufts, attached directly to the hull, reveal the direction and steadiness of the wake in the boundary layer, the inflow to the propellers, and the separation toward the end of the stern. (*Official U S Navy photograph*)

layer along a ship's hull In this layer, motion is imparted to the water in the same direction as the ship Acceleration of the water particles increases as the particles approach the hull. The particles attain ship speed at the interface between the fluid and the hull. The boundary layer generally increases in thickness along the ship from the bow to the stern. At or near the ship's stern, where the potential flow velocity decreases and the pressure increases, the forward velocity of the water im*mediately adjacent to the hull tends* to exceed the ship speed, and the boundary layer thickens rapidly and separates from the hull (Fig 1). As a result, vortices and eddies occur which are left behind by the ship.

Potential wake. This results from motion imparted to the surrounding water by the ship's hull Water particles are first accelerated forward and outward relative to the shi[illegible] -aft opposite to

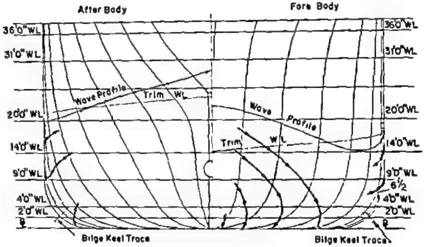


Fig 2 Flow lines around a cargo ship model. The direction of flow is determined primarily by the potential wake. WL = water line (Official U.S. Navy diagram)

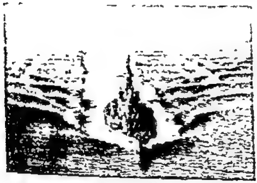


Fig 3 Wave wake. Orbital motions in the transverse wave systems contribute to the subsurface wake (Official U.S. Navy photograph)

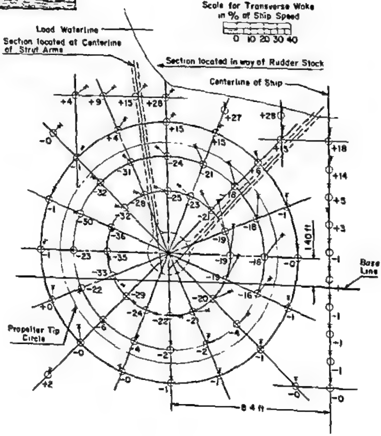


Fig 4 Composite wake diagram for a destroyer model. Measurements were made in the transverse plane of the rudder stock with all appendages fitted and with the propellers rotating (Official U.S. Navy photograph)

Many incentive plans are based on "guesstimates." Increasingly, however, these are being replaced by standards based on work measurement. Newly installed plans are almost always based on work measurement. Similar plans are being installed to pay for extra output by indirect people in maintenance, toolroom, janitor, service and office work. *See* WORK MEASUREMENT.

Wage incentives extend rewards for increased output to productive personnel just as salesmen have long been rewarded for increased sales by commission payments. For salesmen, sales quotas are a measure of standard performance and sales are measured in dollars. Some incentive plans extend throughout a company by being based on profit as a measure of achievement. Personnel may participate in profit sharing plans in proportion to their base pay and longevity with the company. For executives, profit is an appropriate basis for a bonus because it reflects over-all performance. Similarly, many companies reward employees who improve plant performance in ways beyond their direct personal effort by paying a bonus for suggestions that improve methods, equipment, or product; increase rates of output; or reduce costs of materials, supplies, or services.

Except for salesmen, wage incentives have been utilized more with factory people than elsewhere. The prime reason is that during most of our industrial growth, production capacity and sometimes skilled manpower have lagged behind consumer increases. The increased production to meet demands of World Wars I and II provided our most striking examples of greatly accelerated installations of wage incentives.

Average increase in production above past performance reported for 234 installations in World War II was 41.5%; the range was 0 to 103%. The most often quoted past performance is 60% of standard. Usual expectancy of incentive earnings is 25% above standard. These figures suggest that output can be doubled (125/60). Such increases have marked effects on wage costs. They spread overhead expenses and bring down direct labor costs in relation to the difference between past performance and standard output.

Better maintenance, as with equipment, can improve and extend the effectiveness of incentives. In contrast with equipment, however, incentive plans are not impersonal. They relate to people. Emotions, pride, skill and prejudice are mixed with the question of what is a fair day's work. Increases in output, demanded by progress and improvement, are more noticeable with incentives because standards or [illegible] Nevertheless the advanta[illegible]
Wage [illegible] plied skill. This is in accord with our competitive system and our recognition of the individual. (*See* METHODS ENGINEERING.) [P.C.]

Bibliography: P. Carroll, *Better Wage Incentives*, 1957; P. Carroll, *How to Control Production Costs*, 1953.

Wake (seagoing vessels)

The disturbance astern created by a ship in motion. The disturbance is generated in water by the passage of the underwater body of the ship and its propulsion devices, and in air by the above-water body and superstructure. Wake is the composite phenomenon resulting from four principal contributors: (1) boundary layer flow created by the frictional resistance of the hull, (2) potential flow caused by the displacement of the streamlines by the body, (3) standing surface wave patterns caused by gravity effects, and (4) slip streams from propellers.

Frictional wake. This originates from the shearing resistance of a viscous fluid to the passage of a body, and appears as the growth of a boundary

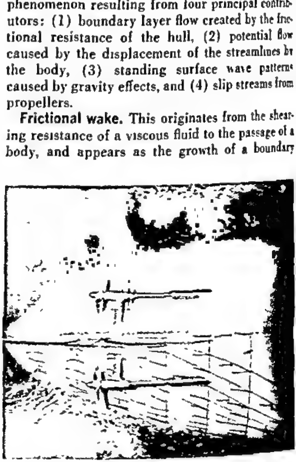


Fig. 1 Wake at the stern of a supertanker model. Nylon tufts, attached directly to the hull, reveal the direction and steadiness of the wake in the boundary layer, the inflow to the propellers, and the separation toward the end of the stern. (*Official U.S. Navy photograph*)

layer along a ship's hull. In this layer, motion is imparted to the water in the same direction as the ship. Acceleration of the water particles increases as the particles approach the hull. The particles attain ship speed at the interface between the fluid and the hull. The boundary layer generally increases in thickness along the ship from the bow to the stern. At or near the ship's stern, where the potential flow velocity decreases and the pressure increases, the forward velocity of the water immediately adjacent to the hull tends to exceed the ship speed, and the boundary layer thickens rapidly and separates from the hull (Fig. 1). As a result, vortices and eddies occur which are left behind by the ship.

Potential wake. This results from motion imparted to the surrounding water by the ship's hull. Water particles are first accelerated forward and outward relative to the ship, [illegible] aft opposite to

the family are slender, sticklike animals, with-t wings, and are so similar in shape and color the twigs and grasses on which they live that ey usually escape detection. A few, including one orth American species, have wings, and some in e tropics are broad and leaflike. Most of the nited States species are only 3–4 in. long, but at ast one is 6 in long. An East Indian walking ick, which may be up to 13 in. long, has the great-st length of any known living insect.

Walking sticks are plant eaters and may occa-ionally defoliate large areas. For the most part he damage they cause is negligible Many people believe that these harmless insects are deadly poi-sonous.

Sexes are usually separate and reproduction bisexual, but parthenogenesis is not uncommon. There is only one generation a year, the eggs being dropped at random. The walking sticks overwinter in the egg See ORTHOPTERA. [J.D B]

Wall construction

To be acceptable an exterior wall of a building must be pleasing in appearance, structurally sound, and durable; must provide insulation and fire resistance; and must offer protection against condensation of water vapor.

Solid and hollow concrete masonry and brick walls are used extensively for low buildings Reinforced concrete walls, precast concrete walls, and precast walls erected by the tilt-up method also offer a variety of surface finishes, low maintenance, strength, and durability.

Plain or corrugated sheets of cement-asbestos and corrugated metal sheets are used for walls of industrial buildings Prefabricated panels consisting of two sheets of metal with a layer of insulation between provide a complete wall section which can be quickly attached to steel framing. Such panels are light in weight and low in heat transmission. See BUILDINGS, CONCRETE, MASONRY. [C N.G]

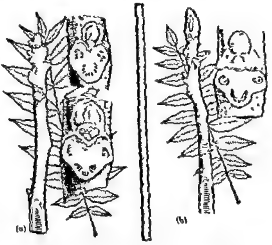
Walleye

Frequently called walleyed pike, pike-perch or yellow pike *Stizostedion vitreum*, is a member of the family Percidae The walleye is one of the most desirable food and game fishes in North American fresh water. Although it ranges southward to Arkansas, Alabama and North Carolina the walleye is most common in the Great Lakes and Upper Mississippi River drainage systems, northward to Great Slave Lake. It is an important commercial fish and is much sought by anglers. Walleyes generally weigh 4 lb or less, but some weigh 10–15 lb

There is a blue subspecies in Lake Erie, usually called the blue pike Formerly in extreme abundance it declined sharply in the early 1950s and is now relatively rare. Closely related to the walleye is the sauger, *Stizostedion canadense*, a smaller, browner fish frequently not distinguished from the walleye by anglers. It occupies about the same territory as the walleye, but prefers somewhat silty rivers and large lakes. It has attained prominence in the sport fishery of the Tennessee Valley and Missouri River artificial lakes. See PERCIFORMES. [J D B.]

Walnut

Black walnut, *Juglans nigra*, a tree of the walnut family. There are two species in North America, both with pinnately compound leaves. See LEAF (BOTANY) The outer husk of the black walnut fruit does not split at maturity, the nut is variously sculptured, the twigs have chambered pith, and the leaves have 15–23 leaflets See FRUIT (BOTANY), PITH. The bark is dark brown to black in color and deeply furrowed (see BARK). The black walnut grows naturally in the eastern half of the United States and in southern Ontario. In deep al-



(a) Black walnut, *Juglans nigra* (b) Butternut, *J. cinerea* (From A H Groves, *Illustrated Guide to Trees and Shrubs*, rev. ed., Harper, 1956)

luvial soils in Maryland, Pennsylvania, Virginia, and west into Iowa and Missouri, specimens have attained 150 ft in height and 6 ft in diameter with clear boles (trunks) of 50–60 ft Trees 100 ft high with a diameter of 3 ft are fairly common The dark brown, coarse-grained wood is comparable in strength to white oak (see OAK) Easily worked, it has long been used for fine furniture and interior paneling. It is also a popular material for gunstocks and airplane propellers because it has strength and shock-resisting ability without undue weight The stand of black walnut in the United States is probably more than 1,500 000,000 board ft The annual cut is 30,000 000–40,000,000 board ft Black walnut is often planted as an ornamental tree.

Butternut, *J. cinerea*, is a tree of medium size which grows mainly in the northeastern quarter of the United States and in adjacent parts of Canada Compared with black walnut, it has lighter colored bark and softer wood The elongated, sticky fruit has a thick, indehiscent husk which encloses a raggedly sculptured nut containing a sweet, oily kernel The three-lobed leaf scars, resembling

the direction of motion, and finally forward and inward at the stern to an area close in behind the hull. Absolute velocity of the water increases and pressure decreases alongside and underneath the middle body of the ship. When viewed from on board ship, these particle motions trace streamlines (see Fig. 2).

Wave wake. This is projected downward from the surface wave patterns of a ship, particularly from the transverse wave system. Water particles below a wave crest are accelerated forward, in the direction of ship and wave motion, and aft below a wave trough. The orbital velocities of the particles decrease exponentially with depth (see Fig. 3).

Propeller wake. This originates from induced velocity forward of the propeller, and then, aft of the propeller, becomes the propeller slip stream impelled aft Rotational as well as axial velocities are generated in the wake by the propeller. Propeller shafts, struts and bossings, as well as other appendages on the hull, leave their own wake patterns.

Wake measurements. These are taken around and astern of a ship, and record the aggregate of the superimposed frictional, potential, wave, and propeller wakes. Wake structure is highly complex and often unsteady Most instruments, such as 5-hole and 13-hole pitot tubes, measure only the average velocity and direction at a point (*see* FLOW MEASUREMENT).

Average wake parallel to the direction of ship travel is called positive when water particle movement is forward and negative when aft. referred to the particles at rest position. Wake fraction w, as defined by D. W. Taylor, is the difference between the ship speed v and the speed of advance of the water v_A at a particular point divided by the ship speed:

$$w = \frac{v - v_A}{v}$$

In wake diagrams (Fig. 4) the number at each point denotes Taylor's wake fraction for the fore and aft component expressed as a per cent. and the vector at each point represents the magnitude of the transverse wake component.

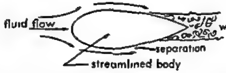
Wake persistence. The "lifetime" of a wake depends upon the wake source. Eddying and currents *resulting from separation of the frictional boundary* layer will continue astern for many ship [illegible]

after several ship lengths. Traces of propeller wake may exceed distances of 50 propeller diameters astern and, according to H. E. Saunders, may even predominate in persistence over the frictional, potential, and wave components of the hull wake. *See* HYDRODYNAMICS; SHIP PROPULSION; WAKE FLOW. [F.A.W.]

Bibliography: S. A. Harvald, *Wake of Merchant Ships*, 1950; H. E. Saunders, *Hydrodynamics in Ship Design*, 1957.

Wake flow

Turbulent eddying flow that occurs downstream from bluff bodies. When fluid flows along the boundary of a solid body, the fluid near the boundary is slowed down by viscous shear stresses exerted at the boundary. This action is progressive along the body, and under certain conditions fluid near the boundary is brought to rest, which causes the fluid moving near the body to separate from the body (see illustration). A wake develops and produces additional form drag on the body.



Wake formed downstream from a streamlined body

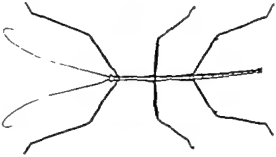
When the flow does not separate from the boundary, as in ideal fluid flow situations, the high fluid velocity at the largest cross section of the body is reduced along the downstream portion of the body with recovery of pressure. Formation of a wake stops the pressure recovery process, leaving a low pressure intensity in the wake and a resultant pressure force on the body acting in the downstream direction. The fluid in the wake is highly turbulent, containing vortices that are shed from the body The wake continues downstream with the flow.

A wake is formed downstream from bluff bodies such as bridge piers, smoke stacks, buildings, or trees. For unsteady flow cases, such as the motion of a train or a ship, the fluid behind the moving object has great turbulence remaining in it, and in the case of a ship is easily discerned for a great distance *See* WAKE (SEAGOING VESSELS); *see also* FLUID-FLOW PRINCIPLES. [V.L.S]

Walking stick

Any member of the insect family Phasmidae, order Orthoptera; also called devil's walking stick.

There are about 100 species of walking stick, mostly tropical. A few species occur in the United States, mainly limited to the South. Most members



The walking stick, Diapheromera femorata; length to 4 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

e family are slender, sticklike animals, with- ings, and are so similar in shape and color e twigs and grasses on which they live that usually escape detection. A few, including one h American species, have wings, and some in tropics are broad and leaflike. Most of the ed States species are only 3–4 in. long, but at one is 6 in long. An East Indian walking , which may be up to 13 in long, has the great- length of any known living insect.

Walking sticks are plant eaters and may occa ally defoliate large areas. For the most part damage they cause is negligible. Many people eve that these harmless insects are deadly poi- ous.

Sexes are usually separate and reproduction exual, but parthenogenesis is not uncommon. ere is only one generation a year, the eggs being opped at random. The walking sticks overwinter the egg. *See* ORTHOPTERA. [J.D.R.]

all construction

o be acceptable an exterior wall of a building ust be pleasing in appearance, structurally ound, and durable; must provide insulation and re resistance, and must offer protection against ondensation of water vapor.

Solid and hollow concrete masonry and brick walls are used extensively for low buildings Rein orced concrete walls, precast concrete walls, and precast walls erected by the tilt-up method also offer a variety of surface finishes, low maintenance strength, and durability.

Plain or corrugated sheets of cement-asbestos and corrugated metal sheets are used for walls of industrial buildings Prefabricated panels consist- ing of two sheets of metal with a layer of insula- tion between provide a complete wall section which can be quickly attached to steel framing Such panels are light in weight and low in heat transmission *See* BUILDINGS, CONCRETE, MA- SONRY. [C.N.C.]

Walleye

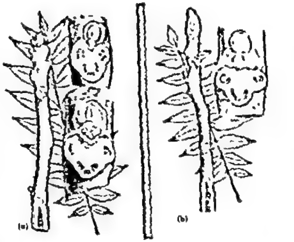
Frequently called walleyed pike, pike-perch or yellow [illegible]

[illegible] water Although it ranges southward to Ar kansas, Alabama and North Carolina, the walleye is most common in the Great Lakes and Upper Mississippi River drainage systems, northward to Great Slave Lake It is an important commercial fish and is much sought by anglers Walleyes gen- erally weigh 4 lb or less, but some weigh 10–15 lb

There is a blue subspecies in Lake Erie, usually called the blue pike Formerly in extreme abun- dance it declined sharply in the early 1950s and is now relatively rare Closely related to the walleye is the sauger, *Stizostedion canadense*, a smaller, browner fish frequently not distinguished from the walleye by anglers It occupies about the same territory as the walleye, but prefers somewhat silty rivers and large lakes. It has attained prominence in the sport fishery of the Tennessee Valley and Missouri River artificial lakes. *See* PERCIFORMES [J.D.R.]

Walnut

Black walnut, *Juglans nigra*, a tree of the walnut family There are two species in North America, both with pinnately compound leaves. *See* LEAF (BOTANY). The outer husk of the black walnut fruit does not split at maturity, the nut is variously sculptured, the twigs have chambered pith, and the leaves have 15–23 leaflets. *See* FRUIT (BOT- ANY); PITH The bark is dark brown to black in color and deeply furrowed (*see* BARK). The black walnut grows naturally in the eastern half of the United States and in southern Ontario In deep al-



(a) Black walnut, Juglans nigra (b) Butternut, J cine- rea (From A H Graves, Illustrated Guide to Trees and Shrubs, rev ed, Harper, 1956)

luvial soils in Maryland, Pennsylvania, Virginia, and west into Iowa and Missouri, specimens have attained 150 ft in height and 6 ft in diameter with clear boles (trunks) of 50–60 ft Trees 100 ft high with a diameter of 3 ft are fairly common. The dark brown, coarse-grained wood is comparable in strength to white oak (*see* OAK) Easily worked, it has long been used for fine furniture and interior paneling It is also a popular material for gun- stocks and airplane propellers because it has strength and shock resisting ability without undue weight The stand of black walnut in the United States is probably more than 1 500,000,000 board ft. The annual cut is 30,000 000 10 000 000 board ft Black walnut is often planted as an ornamental tree.

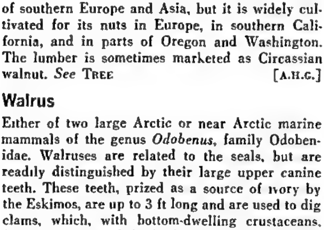
Butternut, *J cinerea* is a tree of medium size which grows mainly in the northeastern quarter of the United States and in adjacent parts of Canada Compared with black walnut, it has lighter-colored bark and softer wood The elongated, sticky fruit has a thick, indehiscent husk which encloses a rag- gedly sculptured nut containing a sweet, oily kernel The three-lobed leaf scars, resembling

those of the black walnut, are surmounted by a short hairy ridge or pad, a feature not present in the black walnut. The diaphragms of the chambered pith of butternut are dark, whereas those of the black walnut are buff-colored. Butternut trees are short lived, being subject to insect and fungus pests (*see* FUNGI; INSECTA). They are more valuable for their nuts and for ornamental purposes than for lumber.

English or Persian walnut, *J. regia*, is a native of southern Europe and Asia, but it is widely cultivated for its nuts in Europe, in southern California, and in parts of Oregon and Washington. The lumber is sometimes marketed as Circassian walnut. *See* TREE [A.H.G.]

Walrus

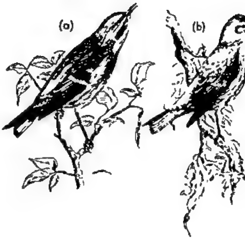
Either of two large Arctic or near Arctic marine mammals of the genus *Odobenus*, family Odobenidae. Walruses are related to the seals, but are readily distinguished by their large upper canine teeth. These teeth, prized as a source of ivory by the Eskimos, are up to 3 ft long and are used to dig clams, which, with bottom-dwelling crustaceans, form the principal food of the walrus. Although most adults are only about 10 ft long, specimens 16 ft long and weighing 3000 lb are known These huge animals usually live in noisy herds. The Eskimos hunt them for their tusks, skins, flesh, and fat, a large walrus yielding almost a ton of oil. *See* CARNIVORA. [J.D.B.]



The walrus, Odobenus rosmarsu; length 10–12 ft (From P. M. Duncan, ed., Cassell's Natural History, Cassell)

Warbler

Any of 109 species of small songbirds, all American, comprising the family Parulidae The warblers, or wood warblers as ornithologists prefer to call them, are a group of active, usually small and slender, perching birds, distinguished from the closely related vireos by their slender bills and more active habits. Although many of the species are dull colored, others are brilliant, and the great variety of color and color patterns makes them second in plumage diversity only to the hummingbirds among all North American birds. There are 54 species in 16 genera in the United States



Warbler. (a) Myrtle, Dendroica coronata, length to 6 in. (b) Yellow, Dendroica petechia; length to 5¼ in (From E L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Typically the warblers are tree-top birds, but there are exceptions here as with every other generality connected with these birds The yellow-breasted chat, for example, is large, thick billed and a brush dweller, but still distinctly a warbler.

Many of the warblers show marked seasonal and sexual dimorphism, and the autumn migrants are difficult to identify, in contrast with the brilliant and distinctly marked spring males. As migrants, many of the warblers move northward rather late in the season, when they are well concealed by the leaves

Warblers are insect eaters, scanning the trees for food with alert, quick movements, hurriedly moving from one spot to another. The ovenbird and water thrush, with olive backs and streaked breasts, *are ground dwellers which look and act* more like thrushes then typical warblers The black and white warbler is a creeper, scanning tree trunks and limbs much in the manner of a nuthatch It is frequently mistaken for a woodpecker

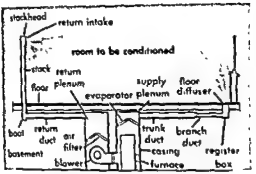
Perhaps best known is the yellow warbler, much like a canary but with red streaks on the breast It is one of the few birds to recognize a cowbird's egg, when it finds one, it will build a false bottom in its nest and start a new clutch of eggs.

Among the showier warblers are the redstarts, the red-faced, blackburnian, chestnut-sided, and black-throated blue warblers.

Most of the warblers winter in South and Central America, but one, the myrtle warbler, spends the winter as far north as the Ohio Valley, Missouri Oregon and sometimes even around Lake Erie and in Massachusetts It further differs from most other warblers in maintaining flocks in the fall and winter It is a summer resident of the northern coniferous forest. *See* PASSERIFORMES [J.D.B.]

Warm air heating system

In a general sense, a warm air heating system is one which circulates warm air. Under this definition both a parlor stove and a steam blast coil circulate warm air. Strictly ... however, a warm air system is one ... iree



Air passage in a warm air duct system (From *Winter Air Conditioning* by Konzo, Carroll, Bareither, The Industrial Press)

furnace surrounded by a bonnet through which air circulates to be heated

When air circulation is obtained by natural gravity action, the system is referred to as a gravity warm air system. If positive air circulation is provided by means of a centrifugal fan (referred to in the industry as a blower), the system is referred to as a forced air heating system

Direct fired furnaces are available for burning of solid liquid, or gaseous fuels, although in recent years oil and gas fuels have been most commonly used Furnaces have also been designed which have air circulating over electrical resistance heaters A completely equipped furnace blower package consists of furnace, burner, bonnet, blower filter, and accessories The furnace shell is usually of welded steel The burner supplies a positively metered rate of fuel and a proportionate amount of air for combustion A casing or jacket encloses the furnace and provides a passage for the air to be circulated over the heated furnace shell The casing is insulated and contains openings to which return air and warm air ducts can be attached The blower circulates air against static pressures, usually less than 1.0 in water gage The air filter removes dust particles from the circulating air The most common filter is composed of 1 to 2-in. thick fibrous matting, although electrostatic precipitators are sometimes used

Accessories to assure effective operation include automatic electrical controls for operation of burner and blower and safety control devices for protection against (1) faulty ignition of burner and (2) excessive air temperatures

Ratings of warm air furnaces are established from tests made in laboratories under industry-specified conditions The tests commonly include heat input rate bonnet capacity and register delivery. Heat-input rate is the heat released inside the furnace by the combustion of fuel in Btu/hr Bonnet capacity refers to the heat transferred to the circulating air, in Btu/hr Register delivery is the estimated heat available at the registers in the room after allowance for heat loss from the ducts has been made in Btu/hr.

The recommended method for selection of a furnace is to ... the total heat loss from the structure under design weather conditions, including the losses through the floor and from the basement, and to choose a furnace whose bonnet capacity rating is equal to, or greater than, the total design heat loss.

The complete forced air heating system consists of the furnace-blower package unit; the return air intake, or grille, together with return-air ducts leading from the grille to the return-air plenum chamber at the furnace, and the supply trunk duct and branch ducts leading to the registers located in the different spaces to be heated

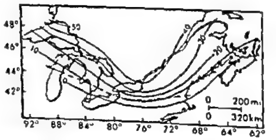
The forced-air system in recent years has no longer been confined to residential installations. The extreme flexibility of the system, as well as the diversity of furnace types, has resulted in widespread use of the forced air furnace installations in the following types of installations, both domestic and commercial: residences with basement, crawl space, or with concrete floor slab, apartment buildings with individual furnaces for each apartment; churches with several furnaces for different zones of the building, commercial buildings with summer winter arrangements; and industrial buildings with individual furnace-duct systems in each zone *See* COMFORT CONTROL, HEATING, COMFORT.

[S.KO.]

Bibliography: S Konzo, J. R Carroll, and H. D. Bareither, *Summer Air Conditioning*, S. Konzo, J R Carroll, and H D Bareither, *Winter Air Conditioning*, National Warm Air Heating and Air Conditioning Assoc., *Heat Gain Calculation and System Design for Summer Air Conditioning*, Manual 11

Warping, earth crust

Gentle bending of the crust of the earth without pronounced folds or dislocations. The earth's crust is being warped at measurable rates believed to be associated with several causes. Measurements have been made by tide gages, tilt meters, repeated leveling and geodetic surveys, and geological observations



Warping illustrated by deformation in eastern North America, calculated from tide-gage records Continuous curves indicate amount of movement (in cm/100 yr) relative to a theoretical datum The broken line marks the outer limit of warping of the Algonquin strand line and its eastward extension (After Gutenberg, in R F Flint, *Glacial and Pleistocene Geology*, Wiley, 1957)

Melting of former ice sheets has led to uplift as in northern Finland, estimated to be 476 m in the past 9950 yr. Uplift is still proceeding there at a maximum rate of 1.3 cm/yr, but sea level is stationary in Denmark (*see* BALTIC SEA). Hudson Bay is likewise rising, and the Great Lakes are tilting southward at 0.1 cm/(km)(yr).

As a result of loading in deltas and sedimentation elsewhere many coasts are sinking; for example, that of Holland and the Atlantic and Gulf coasts of North America at rates of about 0.5 cm/yr, which is accentuated by the general rise of sea level of 0.1 cm/yr. *See* DELTA.

Mountain uplift and faulting are causing other areas such as Japan and the north Pacific Coast of North America to rise at similar average rates with periodic jumps of up to several meters during earthquakes. Relative horizontal movements of the order of 5 cm/yr are observed, for example, in California and New Zealand

Larger tectonic warpings have occurred in the past at unknown rates, as marine sediments high on many mountains and deep in oil wells prove. Drowned beaches and moats, which surround the Hawaiian volcanoes, demonstrate warping due to loading

Many authorities have postulated still larger warping movements due to polar wandering, continental drift, or convection currents in the earth's interior, but the existence, causes, and rates of these are still debated.

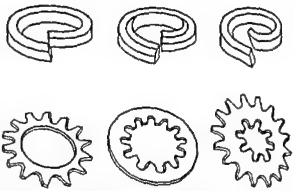
Over large areas and long periods, the earth's crust evidently seeks isostatic equilibrium, but it may be warped by disturbances *See* DIASTROPHISM; OROGENY; TECTONOPHYSICS *See also* SEA LEVEL FLUCTUATIONS; STRAND LINE. [J T W]

Wart

A papillomatous growth which occurs singly or in groups on the skin surface and is often classified according to its location, for example, digital, genital, plantar, or laryngeal It is also called a verruca The agent responsible can be filtered through Berkefeld filters, will produce the disease in human volunteers, and can be seen in the intranuclear inclusions in infected cells by electron microscopy. It is therefore considered to be a virus Only the laryngeal papillomas have been successfully transferred to lower animals. Warts, which spread by contact, respond to many types of treatment or disappear spontaneously. *See* MICROBIOLOGICAL METHODS; VIRUS. [A E M]

Washer

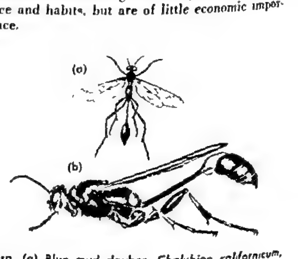
Three types of washer are in common use: plain, spring-lock, and antiturn (tooth-lock washers). Standard plain washers are used to protect a part from damage or to provide for a wider distribution of the load. Because a plain washer will not prevent a nut from turning, a locking type washer should be used to prevent a bolt or nut from loosening under vibration (as illustrated). For industrial applications, spring-lock washers are intended to compensate for possible loosening that can develop between assembled parts and to facilitate assembly and disassembly. Lock washers create a continual pressure between the parts and the fastener. The antiturn type washers may be externally serrated, internally serrated, or both The bent teeth bite into the bearing surface to prevent the nut from turning and the fastening from loosening under vibration. To speed up assembly, a variety of permanent preassembled bolt and washer and nut and washer combinations are available *See* SCREW FASTENER. [W.J.L]



Lock washers. (From W. J. Luzadder, Fundamentals of Engineering Drawing, 4th ed, Prentice-Hall, 1959)

Wasp

Any of a variety of insects, in several families, of the order Hymenoptera. The term wasp has rather loose usage, and generally is used for any hymenopteran with a narrow waist and relatively long, slender body, except the ants. Several wasps are incorrectly called flies, such as the ichneumon and chalcid flies A large number of the wasps are small, parasitic insects and are generally considered useful. At least one, the fig wasp, is a vital pollinator of fig trees. Several species are destructive to crops, notably the wheat straw worm, wheat jointworm and the clover seed chalcid. The large group of gall wasps are of interest to entomologists because of their great variety, both in appearance and habits, but are of little economic importance.



Wasp. (a) Blue mud dauber, *Chalybion californicum*, length ¾ in (b) Yellow mud dauber, *Sceliphron caementarium*; length 1 in. (From E. L. Palmer, Fieldbook of Natural History, McG

Ordinarily, the word wasp refers to members of the families Vespidae, Sphecidae, and their relatives, these are the true social and solitary wasps, hornets, yellow jackets, and mud daubers. These insects are predatory upon other insects and spiders. They usually build nests of paper or mud or excavate nests in the ground. The Sphecidae fill their nests with insects which have been paralyzed by a sting, then deposit their eggs in the nest cells with the food. The Vespidae usually chew up insects with which they feed their larvae. The larvae of all wasps are of the maggot type, white and helpless.

There is almost limitless variation in the details of life histories and nest construction. Many wasps defend their nests fiercely and can inflict painful stings. *See* INSECTA; HYMENOPTERA. [J D B]

Watch

A portable timepiece; commonly a wrist watch, less often a pocket watch. The watch dates from the sixteenth century, when the coiled steel spring was invented. The watch increased in accuracy as longer mainsprings were fabricated and more uniformly operating escapements were invented. The balance wheel with coil hair spring (replacing the straight hog bristle) is commonly used today. Because friction is the principal form of power consumption in a watch, jewels for reduction of friction in the gear and pinion transmission train came into use during the eighteenth century. Traditionally, time is indicated by pointers called hands, but may also be indicated by numbers on a wheel appearing in a window. The mainspring was first wound by a removable key, later by an attached stem, and recently in some watches by an electric watch motor powered by a self-contained dry cell, or by changes in temperature or air pressure. Some watches contain an alarm buzzer. In stop watches, the hands can be abruptly fixed in position, to record the time of occurrence of an event, or reset to zero. *See* ESCAPEMENT; GEAR TRAIN; PAWL; SPRING (MECHANICAL); *see also* CHRONOMETER; CLOCK; HOROLOGY; TIME. [F H R]

Water

The chemical compound with two atoms of hydrogen and one atom of oxygen in each of its molecules. It is formed by the direct reaction of hydrogen with oxygen.

$$2H_2 + O_2 \rightarrow 2H_2O$$

The other compound of hydrogen and oxygen, hydrogen peroxide, readily decomposes to form water.

$$2H_2O_2 \rightarrow 2H_2O + O_2$$

Water also is formed in the combustion of hydrogen-containing compounds, in the pyrolysis of hydrates, and in animal metabolism.

Wa[illegible]

[illegible] late. Water vapor con[illegible] hich move nearly inde[illegible] he relative positions of

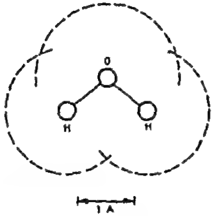


Fig 1 The water molecule

the atoms in a water molecule are shown in Fig. 1. The dotted circles show the effective sizes of the isolated atoms. The atoms are held together in the molecule by chemical bonds which are very polar, the hydrogen end of each bond being electrically positive relative to the oxygen. When two molecules near each other are suitably oriented, the positive hydrogen of one molecule attracts the negative oxygen of the other, and while in this orientation, the repulsion of the like charges is comparatively small. The resulting net attraction is strong enough to hold the molecules together in many [illegible]

1200°C, water vapor dissociates appreciably to form hydrogen atoms and hydroxyl free radicals:

$$H_2O \rightarrow H + OH$$

These products recombine completely to form water when the temperature is lowered. Water vapor also undergoes most of the chemical reactions of liquid water and, at very high concentrations, even shows some of the unusual solvent properties of liquid water. Above 374°C, water vapor may be compressed to any density without liquefying, and at a density as high as 0.4 g/cm³, it can dissolve appreciable quantities of salt. These conditions of high temperature and pressure are found in efficient steam power plants. *See* CRITICAL CONSTANT; HYDROGEN BOND.

Solid water. Ordinary ice consists of water molecules joined together by hydrogen bonds in a regular arrangement, as shown in Fig. 2. The circles represent only the positions of the atoms, but if the sizes, as indicated in Fig. 1, are superimposed upon the figure, then it appears that there is considerable empty space between the molecules. This unusual feature is a result of the strong and directional hydrogen bonds' taking precedence over all other intermolecular forces in determining the structure of the crystal. If the water molecules were rearranged to reduce the amount of empty space, their relative orientations would no longer be so well-suited for hydrogen bonds. This rearrangement can be produced by compressing ice to pressures in

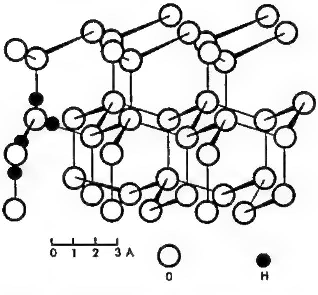


Fig 2. The structure of ice. The hydrogen atoms are omitted for all but two water molecules.

excess of 2000 atm. Altogether five different crystalline forms of solid water have been produced in this way, the form obtained depending upon the final pressure and temperature. They are all more dense than water, and all revert to ordinary ice when the pressure is reduced. *See* CRYSTAL STRUCTURE.

Water in the liquid state. The molecules in liquid water also are held together by hydrogen bonds. When ice melts, many of the hydrogen bonds are broken, and those that remain are not numerous enough to keep the molecules in a regular arrangement. Many of the unusual properties of liquid water may be understood in terms of the hydrogen bonds which remain As water is heated

Properties of water

Freezing point	0°C
Density of ice, 0°C	0 92 g/cm³
Density of water, 0°C	1 00 g/cm³
Heat of fusion	80 cal/g
Boiling point	100°C
Heat of vaporization	540 cal/g
Critical temperature	374°C
Critical pressure	217 atm
Specific electrical conductivity at 25°C	1×10^{-7}/ohm-cm
Dielectric constant, 25°C	78

from 0°C, it contracts until 4° is reached and then begins the expansion which is normally associated with increasing temperature. This phenomenon and the increase in density when ice melts both result from a breaking down of the open, hydrogen-bonded structure as the temperature is raised. The viscosity of water decreases tenfold as the temperature is raised from 0° to 100°C and this also is associated with the decrease of icelike character in the water as the hydrogen bonds are disrupted by increasing thermal agitation. Even at 100°C, the hydrogen bonds influence the properties of water strongly, for it has a high boiling point and a high heat of vaporization compared to other substances of similar molecular weight. *See* LIQUID.

The electrical conductivity of water is at least 1,000,000 times larger than that of most other nonmetallic liquids at room temperature. The current in this case is carried by ions produced by the dissociation of water according to the equation,

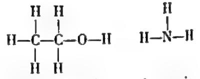
$$H_2O \rightleftharpoons H^+ + OH^-$$

This reaction is reversible, and equilibrium is reached rapidly, so there is a definite concentration of H^+ and OH^- ions in pure water. At 25°C this concentration is 10^{-7} mole/liter of each species or about 10^{14} ions/ml. This concentration of ions is affected by the temperature or by the presence of solutes in the water. *See* ACID AND BASE; HYDROGEN ION.

Pure water, either solid or liquid, is blue if viewed through a thickness of more than 2 meters. The other colors often observed are due to impurities.

Solutions in water. Water is an excellent solvent for many substances, but particularly for those which dissociate to form ions. Its principal scientific and industrial use as a solvent is to furnish a medium for purifying such substances and for carrying out reactions between them. *See* SOLUTION, SOLVENT.

Some substances which dissolve in water with little or no ionization and which are very soluble are



Ethyl alcohol   Ammonia

These are examples of molecules which are able to form hydrogen bonds with water molecules, although, except for the hydrogen of the OH group, it is the hydrogen of the water that makes the hydrogen bond. On the other hand, substances which cannot interact strongly with water, either by ionization or by hydrogen bonding, are only sparingly soluble in it Examples of such substances are benzene, mercury, and phosphorus For discussions of another important class of solutions in water, *see* COLLOID; SURFACE-ACTIVE AGENT.

Chemical properties of water. Water is not a strong oxidizing agent, although it may enhance the oxidizing action of other oxidizing agents, notably oxygen. Examples of the oxidizing action of water itself are its reactions with the alkali and alkaline-earth metals, even in the cold; for instance,

$$Ca + 2H_2O \rightarrow Ca^{++} + 2OH^- + H_2$$

and its reactions with iron and carbon at elevated temperatures,

$$3Fe + 4H_2O \rightarrow Fe_3O_4 + 4H_2$$
$$C + H_2O \rightarrow CO + H_2$$

The second of these reactions is used commercially to produce a gaseous fuel ... lid coke The

gaseous mixture, $CO + H_2$, called water gas, is formed when steam is passed over coke heated to 600°C.

Water is an even poorer reducing agent than oxidizing agent. One of the few substances that *it* reduces rapidly is fluorine, but this reaction is complicated. Chlorine is reduced only very slowly in the cold, according to the equation,

$$2Cl_2 + 2H_2O \rightarrow O_2 + 4H^+ + 4Cl^-$$

An example of another sort of oxidation-reduction reaction in which water plays an essential role beyond that of the solvent is the disproportionation of chlorine,

$$Cl_2 + H_2O \rightarrow HOCl + H^+ + Cl^-$$

which is fast and incomplete in neutral solution but goes to completion if base is added. *See* OXIDATION-REDUCTION.

Substances with strong acid or basic character react with water. For example, calcium oxide, a basic oxide, reacts in a process called the slaking of lime,

$$CaO + H_2O \rightarrow Ca(OH)_2$$

Sulfur trioxide, an acidic oxide, also reacts,

$$SO_3 + H_2O \rightarrow H_2SO_4$$

This reaction occurs in the contact process for the manufacture of sulfuric acid. Both of these reactions evolve enough heat to produce fires or explosions unless precautions are taken.

Another type of substance with strong acidic character is an acid chloride. Two examples and their reactions with water are boron trichloride,

$$BCl_3 + 3H_2O \rightarrow H_3BO_3 + 3HCl$$

and acetyl chloride,

$$CH_3COCl + H_2O \rightarrow CH_3CO_2H + HCl$$

These are often termed hydrolysis reactions, as is the reaction of an ester with water, for instance ethyl acetate,

$$H_3C\overset{\displaystyle O}{\overset{\|}{C}}OC_2H_5 + H_2O \rightarrow H_3C\overset{\displaystyle O}{\overset{\|}{C}}OH + HOC_2H_5$$

A hydrolysis reaction of a different sort is that of calcium carbide used in the production of acetylene

$$CaC_2 + 2H_2O \rightarrow Ca(OH)_2 + C_2H_2$$

For [illegible]

W. [illegible]

solı [illegible]

tact, but in which it *becomes* a part of the structure of the solid. Such compounds are called hydrates, and are formed frequently with the evolution of considerable amounts of heat. Examples range from the hydrates of simple and double salts, calcium chloride hexahydrate, $CaCl_2 \cdot 6H_2O$, and ammonium aluminum alum, $NH_4Al(SO_4)_2 \cdot 12H_2O$, to the gas hydrates which are stable only at low temperatures, for example, chlorine hydrate, $Cl_2 \cdot 6H_2O$, and xenon hydrate, $Xe \cdot 6H_2O$. For further discussion, *see* CLATHRATE COMPOUNDS; HYDRATE.

For various aspects of water, its uses, and occurrence *see* HEAVY WATER; HYDROGEN; HYDROLOGY; IRRIGATION OF CROPS; NUTRITION; OXYGEN; PLANT, WATER RELATIONS OF; PRECIPITATION (METEOROLOGY); SEA WATER; TRIPLE POINT; VAPOR PRESSURE; WATER ANALYSIS; WATER CONSERVATION, WATER MICROBIOLOGY; WATER POLLUTION; WATER POWER; WATER PURIFICATION; WATER SOFTENING, WATER SUPPLY ENGINEERING; WATER TABLE; WATER TREATMENT. [H.L.F.]

Bibliography: N. E. Dorsey, *The Properties of Ordinary Water-Substance in All Its Phases*, 1940; N. V. Sidgwick, *The Chemical Elements and Their Compounds*, 2 vols., 1950

Water analysis

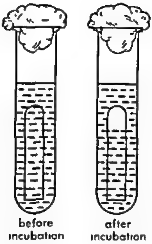
The application of bacteriologic tests to determine the sanitary quality of potable and recreational waters. The objective is detection of animal and human pollution rather than detection of pathogenic bacteria since it may be presumed that water thus polluted is unsafe.

Quantitative testing. The standard plate-count method is used. Measured portions of water are plated with nutrient agar and colonies that develop are counted. Preferably, two sets of plates, incubated at 20°C and 35°C respectively, are used per sample; using a single set, incubation is at 35°C, a temperature favorable for bacteria of human origin. Counts greater than 1000 bacteria per milliliter usually suggest pollution, as does a 35°C count higher than that at 20°C. Quantitative testing is of greatest value in showing deviations from normal conditions in a routinely tested water source. *See* CULTURE TECHNIQUE.

Qualitative testing. This detects particular bacteria or groups of bacteria whose presence indicates fecal pollution. In the United States, the coliform group is almost exclusively employed as the indicator, *Escherichia coli* is used in many other areas (*see* ESCHERICHIA). Coliforms are defined as aerobic and facultative anaerobic, gram-negative, nonsporeforming bacilli which ferment lactose with gas formation, and include *E. coli* and other less common [illegible]

[illegible] highly selective media or environments, like MacConkey broth containing bile salts or the Eijkman test using 45.5°C for incubation, are used for *E. coli* detection. Coliform detection may require the following series of three tests.

Presumptive test. Lactose or lauryl tryptose broth in Durham fermentation tubes is inoculated with the sample and incubated at 35°C. The presence of gas in the inverted vials of the fermentation tubes within 48 hours constitutes a positive result. Absence of gas is interpreted as a coliform-



Durham fermentation tube.

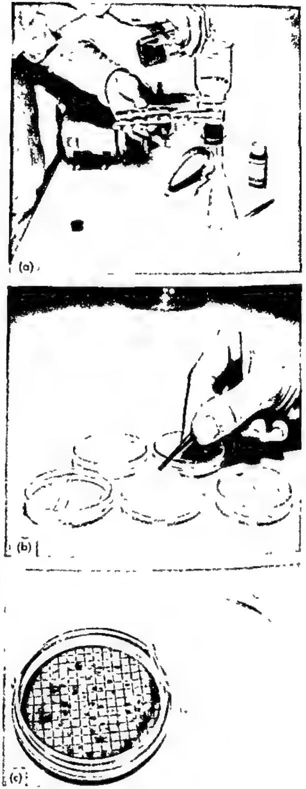
free sample. Gas is presumptive but not conclusive evidence of coliforms since other bacteria, mainly gram-positive forms, may also produce gas. Consequently further testing is required.

Confirmed test. Either liquid or solid media are used. All confirming media are selective, inhibiting bacteria which produce false positive presumptives while permitting coliform growth. Solid media are also differential and make possible recognition of certain coliforms from their colonial morphology. Brilliant green bile broth, the liquid confirming medium, is inoculated from positive presumptive tubes (that is, those tubes which have gas), and the broth is incubated at 35°C. Gas production within 48 hours is interpreted as indicative of the presence of coliforms. Eosin-methylene blue or Endo agar, the solid confirming media, are streaked from positive presumptive tubes and incubated at 35°C for 24 hours. A positive result is the presence of recognizable colonies of coliforms on the plates

Completed test. Since not all coliforms give recognizable colonies on solid confirming media, a completed test is necessary when atypical colonies are found. A transfer is made from an atypical colony to both a nutrient agar slant and a lactose fermentation tube, which are then incubated at 35°C. If colonies of gram-negative nonsporulating rods are found on the slant and gas is produced in lactose broth, coliforms are proven to be present

Most probable number (MPN). Evaluation of water quality is based on the numbers of coliforms present. To determine numbers, replicate portions of a series of decimal dilutions of the sample are inoculated into presumptive tubes. The number of portions and the range of dilutions used vary with the character of the water examined. A statistical analysis of the distribution of positive and negative tubes allows an estimate of the MPN of coliforms in the sample. MPN tables have been prepared that cover the most frequent combinations of sample portions used.

Membrane filters. This alternate technique for coliform detection has been given limited approval by the U.S. Public Health Service. Water is filtered through a cellulosic membrane which collects bacteria on its surface The membrane is placed on a pad saturated with selective medium; the diffusion of nutrients supports development of recognizable coliform colonies on the membrane surface The technique has the advantages of speed, economy of materials and space, and gives a direct rather than statistical estimate of coliform population



(a) Water sample filtered through membrane filter (b) Membrane filter placed on nutrient pad (c) Typical coliform colonies on membrane filter (*Millipore Filter Corp.*)

Sanitary standards. Routine bacteriologic testing cannot provide an al aluation of the

uality of a water. Consequently, quality standards based on MPN determinations are somewhat arbitrary. For potable waters, most communities in the United States follow the U.S. Public Health Service Drinking Water Standards of 1946. These specify the minimum number of samples to be examined each month and the maximum permissible percentage of positive portions, but not specific permissible MPNs. Roughly speaking, to meet these standards, a water must show a MPN of less than 1 per 100 milliliters for most samples analyzed. Swimming pool and recreational water standards are generally less exacting. See MICROBIOLOGICAL METHODS [S.C.R.]

Bibliography: American Public Health Association, *Standard Methods for the Examination of Water, Sewage, and Industrial Wastes*, 1955.

Water boatman

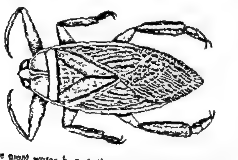
Any member of the insect family Corixidae, order Hemiptera. There are 55 American species. They are up to ¾ in. long, oval, and flattened dorsally. The hind legs are modified for swimming, being long and fringed with hairs. They swim with an oarlike motion. Their front legs are modified for scraping ooze, on which they feed.

The backswimmers of the related family Notonectidae are similar to the water boatmen but swim on their backs, which are deeply keeled. Backswimmers are predaceous and have a powerful beak with which they can inflict a painful bite. See HEMIPTERA [J.D.B.]

Water bug

Any of several large, aquatic insects of the family Belostomatidae, order Hemiptera. Included are the largest species of the order, one United States species being about 3 in. long; a tropical species is twice as large. All United States species are similar in appearance, being brown, broad and flat, with the front legs modified for grasping their prey. They feed upon insects, fish, and amphibians, being able to kill animals four times their size.

These insects frequently leave the water on nocturnal flights of some distance and are attracted by lights; they are often called electric-light bugs. Their bite is quite painful; some caution is required in handling them. The males of some species



The giant water bug, *Lethocerus americanus*, length 3 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

carry the eggs on their backs where they are glued by the female during mating. See HEMIPTERA. [J.D.B.]

Water conservation

The protection, development, and efficient management of water resources for beneficial purposes. Water occupies over 71% of the earth's surface. Its physical and chemical properties make it essential to life and modern civilization. Water is combined with carbon dioxide by green plants in the synthesis of carbohydrates from which all other foods are formed (see PHOTOSYNTHESIS). It is a highly efficient medium for dissolving and transporting nutrients through the soil and throughout the bodies of plants and animals. It can also carry deadly organisms and toxic wastes, including radioactivity. Water is an indispensable raw material for a multitude of domestic and industrial purposes.

Underground water. Water occurs both underground and on the surface. Usable ground water in the United States is estimated to be 47,500,000,000 acre-feet. Annual runoff from the land averages 1,299,000,000 acre-feet. The volume of ground water greatly exceeds that of all fresh water lakes and reservoirs combined. It occurs in several geologic formations (aquifers) and at various depths. Ground water under pressure is known as artesian water, and it may become available either by natural or artificial flowing wells. Ground water, if abundant, may maintain streams and springs during extended dry periods. It originates from precipitation of various ages as determined by measurements of the decay of tritium, a radioisotope of hydrogen found in ground water. The water table is the upper level of saturated ground water accumulation. It may appear at the surface of the earth in marshes, swamps, lakes or streams, or hundreds of feet down. Seeps or springs occur where the contour of the ground intercepts the water table. In seeps the water oozes out, whereas springs have distinct flow. Water tables fluctuate according to the source and extent of recharge areas, the amount and distribution of rainfall, and the rate of extraction. The yield of aquifers depends on the porosity of their materials. The latter represents that portion of water which will drain out by gravity and becomes available by pumping. Shallow ground water (down to 50 ft) is trapped by dug or driven wells, but deep sources require drilled wells. The volume of shallow wells may vary greatly in accordance with fluctuations in rainfall and degree of withdrawal. See GROUND WATER; see also RADIOACTIVE SPECIES PRODUCED BY COSMIC RAYS.

Surface water. Streams supply most of the nation's water needs. Lakes, ponds, swamps, and marshes, like reservoirs, represent stored streamflow. All natural lakes in the United States are calculated to contain 13,000,000,000 acre-feet. Swamps and other wet lands along river deltas, around the borders of interior lakes, and in coastal regions add millions more to the surface supplies. The oceans and salty or brackish sounds, bays,

bayous, or estuaries represent almost unlimited potential fresh water sources. Brackish waters are being used increasingly by industry for cooling and flushing. Reservoirs, dammed lakes, farm ponds, and other small impoundments have a combined usable storage of 300,000,000 acre-feet. The smaller ones furnish water for livestock, irrigation, fire protection, flash-flood protection, fish and waterfowl, and recreation (*see* AGRICULTURE; *see also* separate articles dealing with various aspects of conservation). However, most artificial storage is in reservoirs of over 5000 acre-feet. Lake Mead, located in Arizona and Nevada and formed by Hoover Dam, is the largest (227 square miles) of the 1300 reservoirs, and it contains 10% of the total stored-water capacity, or over 31,000,000 acre-feet These structures regulate streamflow to provide more dependable supplies during dry periods when natural runoff is low and demands are high They store excess waters in wet periods, thus mitigating damaging floods.

Water use. Water withdrawals for all purposes totaled 1,740,000,000,000 gallons per day (1740 bgd) in 1955 in the United States, but actual consumption was only 94 bgd. Exclusive of reservoir evaporation, daily use was 71 bgd, 444 gal per capita or 6% of the nation's runoff—not counted is an additional 20 bgd transpired by western phreatophytes, plants growing along stream banks, canals, and reservoirs (*see* ECOLOGY). Because reservoirs evaporate 23 bgd, interest has developed in underground storage and in coating water surfaces with monomolecular films of oil-derived detergents, such as hexadecanol, to reduce evaporation.

The 115,000,000 people in the United States served by public water supplies in 1955 withdrew 11.3 bgd, and rural dwellers and their livestock used 2 5 bgd Crop irrigation practiced in all 48 states withdrew 110 bgd, the heaviest use being in the 17 drier western states. Only 40% of the irrigation water is returned to streams or aquifers in contrast to 90% of other public water supplies However, pollution prevents reuse of part of the return flow of public water supplies.

Withdrawals for waterpower account for 1500 bgd out of the nation's total, but this use consumes practically none of the water (*see* WATER POWER). Industry takes nearly 116 bgd, mostly for condenser cooling and use in boilers, sanitary services, and cooling machinery in power plants, but industry consumes only 2% of the water utilized Air conditioning averages only 1 1 bgd but may constitute a severe drain on available supplies during hot dry spells in local areas.

Fish, wildlife, and recreation water requirements are nonconsumptive. Clean natural waters and the aquatic environment they create constitute major attractions in undeveloped wilderness areas, national parks, and even in more highly developed agricultural, forest, or suburban localities. However, artificial impoundments, unless properly located, designed, and operated, can destroy or depreciate priceless natural environment.

Water pollution. Streams have traditionally served for waste disposal. Towns and cities, industries, and mines provide thousands of pollution sources. Pollution dilution requires large amounts of water. Treatment at the source is safer and less wasteful than flushing untreated or poorly treated wastes downstream. However, sufficient flows must be released to permit the streams to dilute, assimilate, and carry away the treated effluents (*see* WATER POLLUTION).

Hydrologic cycle. This term refers to the continuous circulation of the earth's moisture. Because of this characteristic, water is considered a renewable natural resource, but underground water is "mined" when it is pumped out faster than its natural renewal and thus may be like a fund resource. The oceans furnish most of the moisture for evaporation and precipitation. *See* PRECIPITATION (METEOROLOGY). Part of the precipitation evaporates, part is returned directly to the oceans, part runs off quickly into watercourses and lakes, part enters the soil or other porous material where it is retained, and the balance enters deep aquifers where it is stored or flows along impermeable underground layers into streams or springs. Solar radiation provides the energy for the hydrologic cycle. Unequal heating of the earth's surface creates air currents of varying temperatures and pressures Warm air masses carry moisture from the oceans. When cooled, the capacity of these air masses to retain moisture is reduced and precipitation results. The source of the air masses and the pressure and temperature gradients aloft determine the form, intensity, and duration of precipitation Precipitation may occur as cyclonic or low-pressure storms (mostly a winter phenomenon responsible for widespread rains), or thunderstorms of high intensity and limited area, or as mountain storms wherein warm air dumps its moisture when lifted and cooled in crossing high land barriers (Fig 1)

Precipitation is characteristically irregular.

The hydrologic cycle. (*Water, USDA Yearbook of Agriculture, 1955*)

Generally, the more humid an area and the nearer the ocean, the more evenly distributed is the rainfall. Whatever the annual average, rainfall in arid regions tends to vary widely from year to year and to fall in a few heavy downpours. Large floods and active erosion result from heavy and prolonged rainfall or rapid melt of large volumes of snow. Flash floods often follow local intense thunderstorms.

Runoff depends on depth, porosity, and compactness of the soil and the underlying material (see HYDROLOGY; see also SOIL), steepness and configuration of the surface, and character and density of the vegetation. Plant crowns, ground cover, litter, and humus dissipate the force of rainfall, thus reducing its power to compact and dislodge mineral soil particles and seal the surface pores. The quantity of water entering the soil during a given time depends on the rate of rainfall

Surface runoff follows when rainfall exceeds the rate at which water is absorbed and transmitted downward. Soil water is retained by adhesion against the pull of gravity in the capillary pores of the soil until their capacity is filled. Such water may evaporate or is available for plant use including photosynthesis and transpiration. Excess moisture in the large pores drains slowly into watercourses or wet weather springs, or enters rock crevices, limestone sinks, and shales, sands, or other permeable materials.

Land management vitally influences the distribution and character of runoff. Inadequate vegetation or surface organic matter, compaction of farm, ranch, or forest soils by heavy vehicles, frequent crop-harvesting operations; repeated burning; or excessive trampling by livestock, deer, or elk all expose the soil to the destructive energy of rainfall or rapid snowmelt. On such lands little water enters the soil, soil particles are dislodged and quickly washed into watercourses, and gullies may form (see LAND USE PLANNING, SOIL CONSERVATION).

Water management problems. These involve economic, social, and intangible values. Efforts to plan and develop river systems for multiple purposes often generate conflicts among different water uses as, for example, irrigation versus navigation on the Missouri River, or hydropower versus salmon or trout fisheries, wildlife, national park, wilderness or historic resources on such rivers as the Columbia, Colorado, or Potomac. Other conflicts stem from actual or threatened dumping of municipal, agricultural (silt), industrial, or acid mine wastes into streams, as occurs upstream from Philadelphia; Washington, D.C.; Cumberland, Maryland, and other cities, or from operations of power dams, irrigation projects, or other uses which restrain or divert flows to the extent of destroying fish habitat or impairing recreational or wildlife resources. Another source of conflict is the mining of ground waters in critically short areas.

Water management technology. This involves the application of biological and engineering principles to attain desired goals. Biological methods for growing upland vegetation having lesser moisture requirements are being studied by the U.S. Forest Service and other agencies (see FOREST AND FORESTRY). Mechanical methods include the practice of waterspreading, which is utilized to desilt floodwaters and to promote the percolation of water into the soil for crop use and ground water recharge. Water which would be a strong pollutant of streams and lakes may be spread on the land. Seabrook Farms in southern New Jersey releases 50 ft per acre of food processing wash water annually onto an 80-acre oak woodland on sandy soil. About 103,000,000 acres of wet lands have been drained in 40 states. Drainage has failed, however, where the suitability of soils for such practice was not adequately determined, or where erosion from adjacent slopes of improperly farmed land silted up the drainage structures. In some instances drainage has drastically reduced waterfowl habitat or aggravated downstream flood damages.

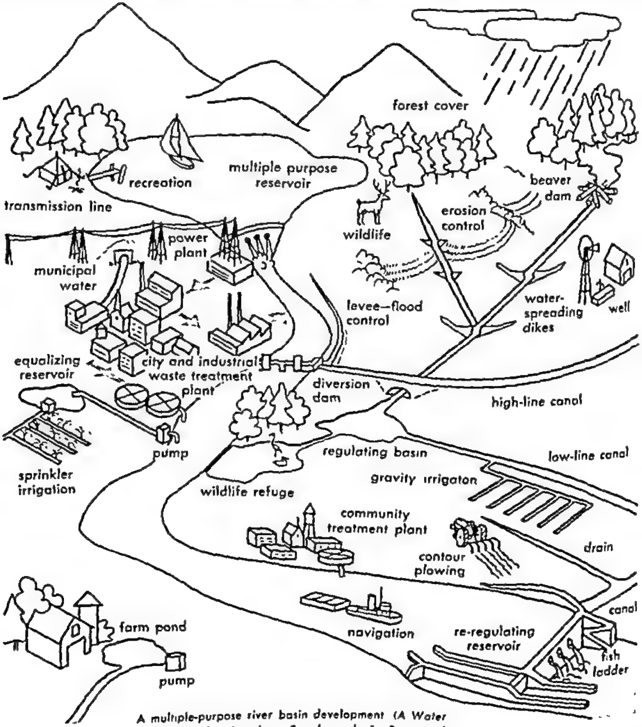
The avoidance of water waste takes several forms. Recycling has permitted huge savings of water, especially in petroleum plants, chemical factories, and steel mills. In some cases reductions have amounted to 96%. Artificial ground-water recharge is successfully practiced in Long Island, New York, where over 311 injection wells conduct water used in air conditioning to underground storage areas for cooling and reuse. A National Association of Manufacturers survey reported that 45% of 3333 industrial establishments applied some kind of water purification treatment. Flows in pipes can be reduced, warmer water can be used, or several grades can be applied by means of separate pipelines. Metering of water stimulates more economical use and encourages repair of leaky connections.

Water rights. Early rights to water followed the riparian doctrine which grants the property owner reasonable use of surface waters flowing past his land unimpaired by upstream landowners. The drier west, however, has favored the appropriation doctrine which advocates the prior right of the person who first applied the water for beneficial purposes, whether or not his land adjoins the stream. Rights to ground water are generally governed by the same doctrines. Both doctrines are currently undergoing intensive study.

State laws generally are designed to protect riparian owners against pollution. States administer the regulatory provisions of their pollution control laws, develop water quality standards and waste-treatment requirements, and supervise construction and maintenance standards of public service water systems. Some states can also regulate ground water use to prevent serious overdrafts. Artesian wells may have to be capped, permits may be required for drilling new wells, or reasonable use demonstrated.

Federal responsibilities consist largely of financial support or other stimulation of state and local water management. Federal legislation permits court action on suits involving interstate streams where states fail to take corrective action following persistent failure of a community or industry to comply with minimum waste treatment requirements. Federal legislation generally requires that benefits of water development projects equal or exceed the costs. It specifies that certain costs be allocated among local beneficiaries but that most of the expense be assumed by the Federal government. In 1955, however, the Presidential Advisory Committee on Water Resources Policy recommended that cost sharing be based on benefits received, and that power, industrial, and municipal water supply beneficiaries pay full cost. These phases of water resource development present difficult and complex questions, because many imponderables enter into the estimates of probable monetary and social benefits from given projects as well as into the cost allocation aspect.

Watershed control. This approach to planning, development, and management rests on the established interdependence of water, land, and people. Conditions on the land are often directly reflected in the behavior of streamflow and in the accumulation of ground water. The integrated approach on smaller watersheds is illustrated by projects under the Watershed Protection and Flood Prevention



A multiple-purpose river basin development (A Water Policy for the American People, vol 1, Report of President's Water Resources Policy Commission, 1950)

Act of 1954 (Public Law 566) as amended by P. L. 1018 in 1956 This act originally applied to floods on the smaller tributaries whose watersheds largely are agricultural, but more recently the application of the act has been broadened to include mixed farm and residential areas. Damages from such frequent floods equal half the national total from all floods. Coordination of structures and land use practices is sought to prevent erosion, promote infiltration, and retard high flows. The Soil Conservation and Forest Services of the U.S Department of Agriculture administer the program The Soil Conservation Service also cooperates with other Federal and state agencies and operates primarily through the more than 2000 soil conservation districts

River basins may be large and complex watersheds For example, the Tennessee River Basin comprises 40,000 square miles in contrast to the 390-square mile upper limit specified in Public Law 1018 Basin projects may involve systems of multipurpose storage reservoirs, intensive programs of watershed protection, and improvement and management of farm, forest, range, and urban lands They may call for scientific research, industrial development, health and educational programs, and financial arrangements to stimulate local initiative The most complete development to date is the Tennessee River Basin, where well planned cooperative activities have encompassed a wide variety of integrated land and water developments, services, and research (Fig 2)

Water conservation organizations. Organizations for meeting water problems take various forms Local or intrastate drainage, irrigation, water supply, or flood control activities may be handled by special districts, soil conservation districts, or multipurpose state conservancy districts with powers to levy assessments Interstate compacts have served limited functions on a regional level To date Congress has not given serious consideration to [illegible]

[illegible] as the Corps of Engineers, Bureau of Reclamation, Fish and Wildlife Service Soil Conservation Service, and Forest Service and to resolve conflicts among agencies and citizen groups Some national and regional civic groups, such as the League of Women Voters, are currently studying alternative approaches to the administration of river basin programs.

International agreements. Cooperation in the control, allocation, and utilization of international waters is authorized by treaties with Canada and Mexico Permanent commissions have been established to deal with specific streams such as the Rio Grande and Colora[illegible] ecally, as [illegible] United Na[illegible] [illegible] Assistance Program and through regional commissions, is promoting cooperative studies and developments [illegible] veloped nations having common boundaries. See separate articles for other important aspects of conservation. [B F.]

Bibliography: See CONSERVATION OF RESOURCES.

Water flea

Any member of the order Cladocera, class Crustacea, phylum Arthropoda. There are several hundred species of Cladocera, most of them freshwater forms, although a few are marine. Most common and best known is the genus *Daphnia*. Sometimes the term water flea is used in a very broad sense, with general reference to all the subclass Entomostraca, or small aquatic crustaceans.

Water fleas are world wide in their distribution; some species are local but many are completely cosmopolitan in the proper environment. They are extremely valuable in the aquatic food chain, forming the principal food for many young freshwater fishes, and for many species of smaller fishes throughout their life. *Daphnia* are reared in tanks and fed to young game fishes in hatcheries, and to tropical fishes of all ages. They are marketed as tropical fish food in both the living and dried condition

The Cladocera are distinguished from other crustaceans by their short, compact bodies; faint segmentation, often enclosed in a bivalve carapace, four to six pairs of thoracic appendages, and the second pair of antennae very large, the first pair reduced The second pair of antennae is biramous and used for swimming, producing a characteristic jerky motion in the water The abdomen is small and usually bent under the thorax. The single eye is compound and median Eggs are developed within a large dorsal brood sac of the female; the young have the form of the adult. In many species there is a winter egg enclosed in a shell made of two curved plates, the ephippium, one, two, or several eggs may be enclosed in a single ephippium. The posterior end of the carapace often ends in a prominent spine.

Water fleas are primarily parthenogenetic. Only females hatch from the winter eggs and these continue to produce parthenogenetic individuals throughout the season Usually 10–20 eggs are produced at a time and retained within the brood pouch until they hatch; numbers may vary from 2 to 40 Males appear in small numbers throughout the season and may at times outnumber the females Male production is thought to occur when there is a shortage of food, crowding of the females, or water temperatures between 14 and 17°C Mating produces sexual eggs, known as winter eggs

Daphnia is typical of the order. Other well known genera include *Cyclops*, significant not only as a food organism but also as an intermediate host in some parasitic worm life cycles, *Diaptomus*, quite similar to *Cyclops*; and the large long and slender (up to 18 mm long) *Leptodira* There are many other genera. See CRUSTACEA; CLADOCERA

[J.D B.]

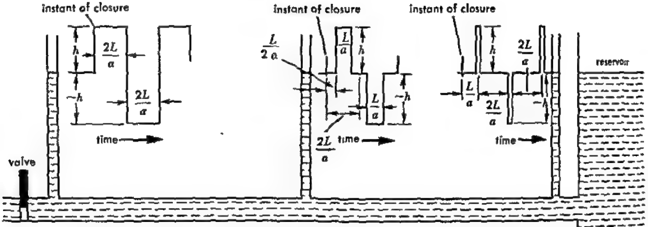


Fig 1. Surge pressures along a conduit for instantaneous closure. (Am. Water Works Assoc.)

Water hammer

An abrupt change of fluid flow in a closed conduit that produces pressure changes or surges. The term water hammer probably developed as a result of the audible and mechanical slamming of valves. The term surge is used in the petroleum pipeline practice and for fluids other than water.

The fundamental phenomenon is the same irrespective of the fluid in the conduit. The basic theory applies equally well to crude oil, gasoline, or water and involves the physical characteristics of the conduit and the fluid. N. Joukovsky and Lorenzo Allievi evolved the elastic wave theory for water hammer in contrast to previous treatments which considered the fluid column as a rigid body or assumed a purely mass action. Joukovsky also demonstrated that water-hammer pressures were direct functions of the abrupt change in fluid velocity and the surge wave velocity in the system independent of the length of the conduit. He also developed the relation for critical time of a conduit as the time required for a surge wave to travel from the point of origin to the point of relief and return. The velocity of the surge wave is that of sound in the system, and is dependent upon the diameter, thickness, and elasticity of the conduit and the specific gravity and compressibility of the fluid.

In the simplest case, consider water flowing in a closed conduit from a reservoir with a valve at the far end.

Figure 1 shows the pressure-time relationship for water hammer at three points in the conduit when the valve is closed in zero time. The pressure rise h is obtained from Eq. (1) in the accompanying list of basic equations.

The elastic effects of the water and pipe walls affect the conditions, and all of the water will not be uniformly retarded. The closure will initially affect only the portions of the flow near the valve. These portions will be compressed and the adjacent walls expanded by the pressure caused by the closure (Fig. 2). The increased pressure will require some time to travel along the pipe before the effect can extend to distant portions of the flow. This rate of travel is determined by the surge wave velocity a, Eq. (2), and the critical water-hammer time for the conduit is given by Eq. (3).

Any change in flow taking place in $\mu = 2L/a$ sec or less will produce an instantaneous water hammer given by Eq. (1).

L. Allievi developed a major extension of elastic wave theory to cover changes in flow taking place in times greater than $2L/a$ sec by computing the net surge pressure change as the algebraic sum of a series of positive and negative surge waves as given in Eq. (4). Charts based on linear changes of flow with respect to time have been developed in many forms. Figure 3 shows a simplified chart for linear flow changes based on equivalent times T_E as developed in Fig. 4.

The theory can also be applied to nonuniform rates of changes in flow using mathematical procedures, arithmetic integration, or graphical methods. Certain problems also lend themselves to the use of electronic or other forms of computers. Changes in diameter, thickness, or materials of construction, as well as partial reflections at branches, can be accounted for in such precise computations of surges. Variations in flow in hydraulic turbines and the behavior of pumps can also be included in analyses.

The extension of water-hammer theory to other fluids such as crude petroleum or refined products is possible when the physical characteristics of the fluids are known. The density and bulk modulus of compressibility are essential. The latter is equally important because it varies with temperature and pressure and for different fluids.

Approximate values in pounds per square inch of bulk modulus for common fluids are 300,000 for water, 230,000 for crude oil, and 130,000 for gasoline.

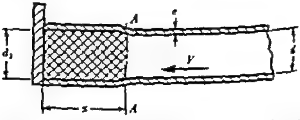


Fig. 2. Example of expansion of conduit walls and fluid compression because of [illegible]. (ASME)

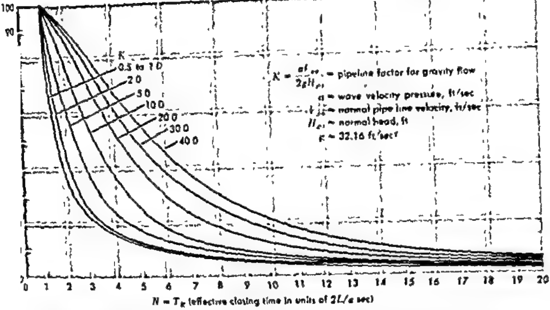
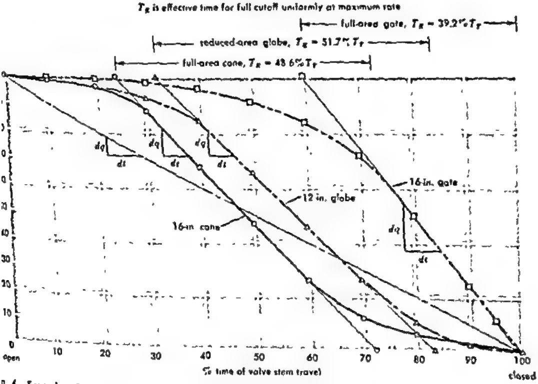


Chart for determination of water hammer pressure. (R S Quick, Comparison and limitations of various water hammer theories, Mechanical Engineering, vol 49, no. 5a)



g 4 Examples of nonuniform rates of flow change for different types of valves (Am. Water Works Assoc.)

The Committee on Water Hammer, American Society of Mechanical Engineers, formed in 1931, sponsored three symposiums and many technical articles on the subject The first symposium in 1933 has been published as a separate volume Approximate formulas from handbooks give results 60–530% of true values and should be avoided.

Many field tests have confirmed the elastic wave theory in all its aspects

Computations, considering all the basic factors,

DRODYNAMICS.

BASIC EQUATIONS FOR WATER HAMMER

Pressure change in critical time or less

$$h = (a/g)V \quad (1)$$

Surge wave velocity

$$a = \frac{12}{\sqrt{\frac{w}{g}\left(\frac{1}{k} + \frac{d}{Ee}\right)}} \quad (2)$$

Critical time of conduit

$$\mu = 2L/a \quad (3)$$

General integrals of basic formulas for all surge action

$$H_{xt} = H_{xo} + F\left(t - \frac{x}{a}\right) + f\left(t + \frac{x}{a}\right) \quad (4a)$$

$$V_{xt} = V_{xo} - \frac{g}{a}\left[F\left(t - \frac{x}{a}\right) - f\left(t + \frac{x}{a}\right)\right] \quad (4b)$$

Time of movement of control devices

$$T_E = T_T(f_v) = N\mu \quad (5)$$

Pipeline factor

$$K = aV_{xo}/2gH_{xo} \quad (6)$$

Maximum rate of change in flow

$$dv/dt \quad (7)$$

where a = surge wave velocity, ft/sec
d = inside diameter of conduit, in
e = thickness of conduit wall, in
E = Young's modulus of elasticity of conduit wall material, lb/in²
f = subnormal pressure wave
f_v = dimensionless coefficient for control device to relate nonuniform change in flow to maximum rate of change in flow, dv/dt
F = supernormal pressure wave
g = gravitational constant, 32.16 ft/sec²
h = pressure change, ft of fluid
H_{xo} = normal pressure
H_{xt} = total pressure at time t
k = bulk modulus of fluid, lb/in² (compressibility factor)
K = pipeline factor
L = length of conduit, ft
N = number of intervals of $2L/a$ sec
t = time, sec
T_E = effective travel time of control device, sec
T_T = total travel time of control device, sec
v = fluid velocity at any time t
V = fluid velocity change, ft/sec
V_{xo} = initial fluid velocity
V_{xt} = total velocity at time t
w = weight of liquid, lb/ft³
x = any point along conduit
μ = critical time, sec

[S L K]

Bibliography: L. Allievi, *Theory of Waterhammer*, trans. E. E. Halmos, 2 vols., 1925; ASME Committee on Water Hammer, First Symposium on water hammer, 1933, reprint, 1949, Second symposium on water hammer, 1937, *Trans. ASME*, 59:651, 1937, Third symposium on water hammer, 1955, *Trans. ASME*, 78:1297, 1956; S. L. Kerr, Effect of valve action on water hammer, *Jour. Am. Water Works Assoc.*, vol. 51, 1959; S. L. Kerr, L. H. Kessler, and M. B. Gamet, New method for bulk-modulus determination, *Trans. ASME*, 72:1143, 1950; J. Parmakian, *Waterhammer Analysis*, 1955, G. R. Rich, *Hydraulic Transients*, 1951; O. Simin, Water hammer, *Proc. Am. Water Works Assoc.*, 24:336, 1904.

Water microbiology

An aspect of microbiology that deals with the normal and adventitious microflora of natural and artificial water bodies. Two major lines of investigation have been pursued, the characterization of the normal flora and its activities, and the public health aspects of potable and recreational waters. See WATER ANALYSIS; WATER-BORNE DISEASE.

Qualitatively, the normal microbial flora is controlled by such factors as temperature, salinity, and chemical composition of the water, and is largely independent of geographical location. Thus, different types of water have highly characteristic populations. For example, thermal springs the world over contain the same groups of blue-green algae, mainly of the family Chroococcaceae and the genus *Phormidium*. Highly saline waters are characterized by the presence of *Dunaliella*, a flagellated green alga. Sulfureted waters contain a complex assemblage of bacteria in whose metabolism sulfur compounds play a special metabolic role.

The numbers of bacteria in unpolluted waters are low. Samples collected in the oceans or large lakes remote from land generally give viable counts of less than 1000 per milliliter; direct microscopic counts, which include both viable and nonviable bacteria, are higher. There is some variation in number of bacteria with depth, distance from land, and season. In contrast, sediments have bacterial populations equal in magnitude to those of soil. The populations of microscopic algae show large seasonal variations and are greatly influenced by the supply of available nutrients. They are limited in distribution to the zone where light is available for photosynthesis, the euphotic zone. At times microscopic forms occur in such tremendous numbers that marked discoloration of the water results, as in the Red Tides of certain coastal waters. See MARINE MICROBIOLOGY.

Surface waters may receive large numbers of bacteria, including pathogens, by runoff from the land surfaces, and by pollution from human and animal sources. Bacteria thus introduced are eliminated quite rapidly either by failure to grow in an unfavorable environment, or by such processes as sedimentation and predation. Thus a natural water not subject to sufficient pollution to change its character maintains its indigenous flora; this fact permits the deliberate disposal of wastes into a water body.

The bacterial flora carries out the same types of biochemical processes in water and sediments as it does on land and is thus responsible for mineralization of organic matter and completion of the cycle of nutrients. In contrast to the situation on land, the microscopic algae are the primary producers of organic matter in the water. See MICROBIOLOGY; WATER PURIFICATION. [C B N]

Water pollution

The impairment of the quality of natural water, interfering with its proper and beneficial uses. Pollution is caused by the various substances in domestic and industrial wastes. It may be classified according to the nature of these substances—physical, chemical and biological. For a discussion of water pollution and natural purification, *see* SEWAGE DISPOSAL. [D.J.O.]

Water power

Power that can be developed from movement of masses of water. Such movement is of two different kinds: (1) the falling of streams of water through the force of gravity and (2) the rise and fall of tides through lunar (and solar) gravitation. Only 0.33% of solar energy expended on the hydrosphere is used to lift water vapor above the earth, but assuming that precipitation falls an average of 10,000 ft, the falling of rain and snow continuously liberates energy at the rate of 600,000,000,000 hp. The only part of this energy that is practically recoverable is the potential energy of water flowing in large stream [illegible] ch [illegible] ten [illegible] about [illegible] of the potential, or about 7×10^{12} kw-hr/year (9.3×10^{12} hp-hr/year), but conceivable improvements in technology might double this figure. The capacity of world water power plants actually installed by the end of 1957 was around 90,000,000 hp. Assuming a world wide use factor of 50% and an over-all efficiency of 80%, hydraulic plants produced in 1957 about 4×10^{11} hp-hr of energy (about 1% of the world's energy requirements).

The installed electric generator capacity in the United States is about 30,000,000 kw, and about 140,000,000,000 kw-hr of electricity is generated annually.

Power from streams. While it is indicated that the practical potential might be 20 times 1957 production, this is little more than a guess because many streams in Africa, Asia, and South America have not been accurately evaluated. The North American continent possesses not quite 13% of the estimated world potential water power, but has about 40% of the total actually installed power. Europe with 10% of the world's potential also has 40% of the total installed power. Thus, the principal water power development of the future can be expected in Africa, Asia, South America, and island countries which, together, have around 77% of the world's potential.

The three Pacific coast states have 34% of the total potential of the United States, and the mountain states, 24%. Thus, the 11 western states with only 14% of the population have nearly three-fifths of the water-power potential. United States water power is almost 20% developed in contrast with only 5% for the world as a whole. In absolute value, water power will gradually increase in the United States, but hydroelectric power has been declining and will continue to decline in proportion to total production of electric power in the United States. The reasons for this are the rapid increase in United States demand for electric power and the decreasingly attractive sites for future hydroelectric development. Not many sites remain where water can be impounded with substantial head. Of the 810 commercial water-power plants in the United States, only 20 have heads of less than 10 ft. A third of the plants have heads of 100 ft or more. The average head is 89 ft. In 1957, less than 30% of United States electric power was hydroelectric in origin.

Installed capacity of hydroplants cannot be counted upon for perpetuity because of the gradual filling of reservoirs with sediment. This effect is serious for irrigation, flood control, and navigation but, on the average, the final result would be a reduction in power capacity of only about 25%. Because the Columbia River carries little silt, the reservoirs at Grand Coulee and Bonneville should last more than a 1000 years, but the Colorado River is so muddy that, in its first 13 years, Lake Mead, behind Hoover Dam, lost 4.5% of its capacity. The most rapid silting of Federal projects has been in the Guernsey Reservoir in Wyoming which in 20 years has lost about one third of its capacity.

Power from tides. A portion of the kinetic energy of the rotation of the earth appears as ocean tides. In total amount, tide energy could supply about 2×10^{13} hp-hr/year, about half the world's needs in 1957, but the practical prospect is for the eventual utilization of not more than 0.3% of the potential. Numerous ingenious schemes have been devised for capturing tide energy: lifting a weight, compressing air, moving paddle wheels by the tide stream, and the filling of basins during flow tide and operating turbines with the difference in head on ebb tide (or vice versa). The basin plan is the most practical. The single-basin plan can not give firm power, since there would be power peaks of varying heights, and periodic absences of power. Also, the times of these peaks and absences would shift from day-to-day. The two-basin plan, with high level and [illegible] ower but at [illegible] a large difference between high and low tides (preferably at least 20 ft) and a shore line of suitable geography for establishing storage basins. The only sites in the continental United States meeting these specifications even in part are the Hudson River, the Maine shore above Penobscot Bay, and certain locations in Alaska. The only tide-power sites that have received serious attention are the Severn River in southern England, the Rance River and Mont St Michel in northern France, the San José and Deseado rivers of Argentina, the Petitcodiac and Memramcook estuaries in the Bay of Fundy, Canada, and Passamaquoddy where Maine joins New Brunswick. No firm power could be available from the Severn project without increasing cost

BASIC EQUATIONS FOR WATER HAMMER

Pressure change in critical time or less

$$h = (a/g)V \tag{1}$$

Surge wave velocity

$$a = \frac{12}{\sqrt{\frac{w}{g}\left(\frac{1}{k} + \frac{d}{Ee}\right)}} \tag{2}$$

Critical time of conduit

$$\mu = 2L/a \tag{3}$$

General integrals of basic formulas for all surge action

$$H_{xt} = H_{xo} + F\left(t - \frac{x}{a}\right) + f\left(t + \frac{x}{a}\right) \tag{4a}$$

$$V_{xt} = V_{xo} - \frac{g}{a}\left[F\left(t - \frac{x}{a}\right) - f\left(t + \frac{x}{a}\right)\right] \tag{4b}$$

Time of movement of control devices

$$T_E = T_T(f_v) = N\mu \tag{5}$$

Pipeline factor

$$K = aV_{xo}/2gH_{xo} \tag{6}$$

Maximum rate of change in flow

$$dv/dt \tag{7}$$

where a = surge wave velocity, ft/sec
d = inside diameter of conduit, in
e = thickness of conduit wall, in
E = Young's modulus of elasticity of conduit wall material, lb/in^2
f = subnormal pressure wave
f_v = dimensionless coefficient for control device to relate nonuniform change in flow to maximum rate of change in flow, dv/dt
F = supernormal pressure wave
g = gravitational constant, 32.16 ft/sec^2
h = pressure change, ft of fluid
H_{xo} = normal pressure
H_{xt} = total pressure at time t
k = bulk modulus of fluid, lb/in^2 (compressibility factor)
K = pipeline factor
L = length of conduit, ft
N = number of intervals of $2L/a$ sec
t = time, sec
T_E = effective travel time of control device, sec
T_T = total travel time of control device, sec
v = fluid velocity at any time t
V = fluid velocity change, ft/sec
V_{xo} = initial fluid velocity
V_{xt} = total velocity at time t
w = weight of liquid, lb/ft^3
x = any point along conduit
μ = critical time, sec

[S.L.K.]

Bibliography: L. Allievi, *Theory of Waterhammer*, trans. E. E. Halmos, 2 vols., 1925; ASME Committee on Water Hammer, First Symposium on water hammer, 1933, reprint, 1949, Second symposium on water hammer, 1937, *Trans. ASME*, 59:651, 1937, Third symposium on water hammer, 1955, *Trans. ASME*, 78:1297, 1956; S. L. Kerr, Effect of value action on water hammer, *Jour. Am. Water Works Assoc.*, vol. 51, 1959; S. L. Kerr, L. H. Kessler, and M. B. Gamet, New method for bulk-modulus determination, *Trans. ASME*, 72:1143, 1950; J. Parmakian, *Waterhammer Analysis*, 1955; G. R. Rich, *Hydraulic Transients*, 1951; O. Simin, Water hammer, *Proc. Am. Water Works Assoc.*, 24:336, 1904.

Water microbiology

An aspect of microbiology that deals with the normal and adventitious microflora of natural and artificial water bodies. Two major lines of investigation have been pursued, the characterization of the normal flora and its activities, and the public health aspects of potable and recreational waters. *See* WATER ANALYSIS; WATER-BORNE DISEASE.

Qualitatively, the normal microbial flora is controlled by such factors as temperature, salinity, and chemical composition of the water, and is largely independent of geographical location. Thus, different types of water have highly characteristic populations. For example, thermal springs the world over contain the same groups of blue-green algae, mainly of the family Chroococcaceae and the genus *Phormidium*. Highly saline waters are characterized by the presence of *Dunaliella*, a flagellated green alga. Sulfureted waters contain a complex assemblage of bacteria in whose metabolism sulfur compounds play a special metabolic role.

The numbers of bacteria in unpolluted waters are low. Samples collected in the oceans or large lakes remote from land generally give viable counts of less than 1000 per milliliter; direct microscopic counts, which include both viable and nonviable bacteria, are higher. There is some variation in number of bacteria with depth, distance from land, and season. In contrast, sediments have bacterial populations equal in magnitude to those of soil. The populations of microscopic algae show large seasonal variations and are greatly influenced by the supply of available nutrients. They are limited in distribution to the zone where light is available for photosynthesis, the euphotic zone. At times microscopic forms occur in such tremendous numbers that marked discoloration of the water results, as in the Red Tides of certain coastal waters. *See* MARINE MICROBIOLOGY.

Surface waters may receive large numbers of bacteria, including pathogens, by runoff from the land surfaces, and by pollution from human and animal sources. Bacteria thus introduced are eliminated quite rapidly either by failure to grow in an unfavorable environment, or by such processes as sedimentation and predation. Thus a natural water not subject to sufficient pollution to change its character maintains its indigenous flora; this fact permits the deliberate disposal of wastes into a water body.

The bacterial flora carries out the same types of biochemical processes in water and sediments as it does on land and is thus responsible for mineralization of organic matter and completion of the cycle of nutrients. In contrast to the situation on land, the microscopic algae are the primary producers of organic matter in the water. *See* MICROBIOLOGY; WATER PURIFICATION. [S.C.R.]

Water purification

The treatment of potable water to destroy disease-producing agents. Purification is employed in all but economically underdeveloped areas and, where practiced, has reduced water-borne diseases to the vanishing point. Chemical disinfection, either by itself or in combination with storage, sedimentation, or filtraton, is the universal practice. Chlorine is the most economical agent for disinfection and is used in most water systems. Disinfection by physical agents is possible but cost considerations have limited the application of such procedures. Boiling is the method of choice for treating small quantities of water of dubious quality for individual or household use. Ultraviolet irradiation has been employed for the disinfection of swimming pool waters and for the treatment of military and a few municipal water supplies where economic factors are not of primary significance. The increasing availability of radioactive wastes from the atomic energy industry renders it theoretically feasible to disinfect water and sewage by shorter wave length irradiations, but much engineering research will be required before practical application is achieved.

Chlorination has as its main purpose the destruction of all pathogenic microorganisms. Secondary benefits may also be derived such as the reduction or elimination of undesirable tastes and odors. Chlorine combines with water to form hypochlorous acid and hypochlorite ion, the proportions of each depending on the pH. Chlorine in these forms, known as free available chlorine, has marked germicidal powers because of its ability to combine with or oxidize classes of organic compounds essential to life. The exact mechanism of killing is still debatable but most probably involves destruction of vital enzymes or structural proteins of the microbes. Chlorine also combines with ammonia and organic nitrogen compounds to form chloramines and chloro derivatives. In these forms, it is known as combined available chlorine, and its germicidal powers are retained, but to a much lesser degree than those of free [illegible] taining some [illegible] dition of su[illegible] first in a co[illegible] available residual chlorine and then, as oxidation goes towards completion, in a free available residual chlorine. For any dosage the difference between the chlorine added and the residual available, combined or free, at an arbitrarily specified time is known as the chlorine demand.

The amount of chlorine required to ensure a safe water will depend on many factors, the most important being pH and temperature of the water, the amounts and types of organic and inorganic matter present, the contact time available, and whether a free or combined residual chlorine will be maintained. Proposed safe minima for free available chlorine after 10 min of treatment range from 0.2 to 0.8 parts per million (ppm) and for combined available after 60 min, 1.0 to 1.8 ppm over a pH range of 6.0 to 9.0. Although these levels are probably adequate for destruction of pathogenic bacteria, higher residuals are required to destroy the cysts of the amebic dysentery organism.

Chlorine, handled in liquid form in tanks under pressure, is introduced into the water by two basic methods: (1) direct feed in which gaseous chlorine is discharged through diffusers directly into the receiving water, and (2) solution feed in which it is first dissolved in a stream of make-up water and then discharged at the desired point in solution. Control of chlorine addition varies from manual to completely automatic systems in which the feed is regulated by the results of chlorination. Operation of a water supply may involve either plain chlorination, which is the addition of chlorine to an otherwise untreated water as it enters the distribution system, or pre-, post-, or rechlorination, where chlorine is added before or after other treating procedures.

Hypochlorination employs calcium or sodium hypochlorite as the disinfectant. The former is available in solid form with at least 70% available chlorine, the latter as a solution of about 15% available chlorine. Although the equipment required for handling calcium hypochlorite is less expensive than that for free chlorine, the cost of the chemical is greater. Consequently, except for emergency uses, hypochlorination is employed only where water consumption is small and the chemical cost is of no significance. Tablets containing calcium hypochlorite in amounts adequate for the treatment of canteen-size quantities of water were issued for individual use by the military services during World War II.

Ozone treatment is employed by a limited number of public water supply systems, mainly in France. Ozone is generated by the discharge of high-voltage electricity through dry air between stationary electrodes. The water is sprayed into an atmosphere of ozone, or the ozonized air is discharged into the water in a mixing chamber. See CHLORINE; SANITARY ENGINEERING; WATER-BORNE DISEASE. [S.C.R.]

Bibliography. American Water Works Association, *Water Quality and Treatment*, 2d ed., 1950.

Water snake

Any of about 100 different species of the genus *Natrix*, found over much of the Northern Hemisphere and in North Africa and Australia. They are moderately large, heavy-bodied snakes, with strongly keeled scales. Currently nine different species are recognized in the United States, some with several subspecies, some of these are frequently considered as distinct species. Although the water snakes are harmless, the general similarity of some to the poisonous cottonmouth causes many people to believe that these are deadly animals. *Natrix* snakes feed on fishes and frogs and are of little economic importance although fisher-

men sometimes believe they are detrimental to sport fishing. They produce large litters of living young. They are active either by day or night. *See* REPTILIA. [J.D.B.]

Water softening

A water-treatment process by which undesirable cations (of calcium and magnesium) are removed from hard waters. The presence of these cations in water is undesirable for household purposes, boiler feed, food processing, and chemical processing, because of reactions that form soap scum, boiler scale, and unwanted by-products. *See* WATER TREATMENT.

Hard waters may be softened by precipitation processes, cation-exchange processes, or combinations of these. The choice of a process depends on a number of factors, among which are the composition of the hard water, the end uses of the softened water, the type or types of hardnesses to be removed, the degree of removal required, and the relative processing costs. In the following discussion, hardness is expressed as calcium carbonate equivalents in parts per million (ppm), and lime means hydrated lime

Cold lime-soda processes. The equipment used in these processes consists of chemical feeders, a softener unit, and usually filters. Chemicals used are lime (or lime and soda ash) plus a coagulant Precipitates produced are calcium carbonate and magnesium hydroxide. Lime may be used to reduce the calcium carbonate hardness to about 35 ppm or to reduce both calcium and magnesium carbonate hardness to about 70 ppm. Lime and soda ash will reduce total hardness (carbonate and noncarbonate) to about 70 ppm without using excess chemicals. With excess chemicals, total hardness may be reduced to 16 ppm.

The cold lime-soda process may be combined with the zeolite process in a two-stage operation in which the filtered effluent from a cold lime-soda softener is passed through a sodium cation-exchanger (zeolite) water softener Either lime or lime and soda ash may be used in the first stage. In either case, the residual hardness from most hard waters is reduced to practically "zero" hardness (1–2 ppm) by the second stage.

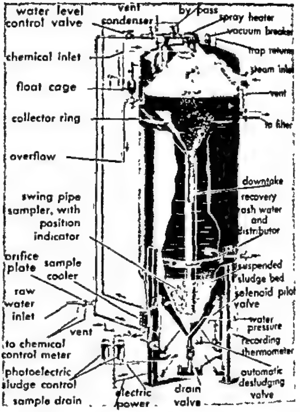
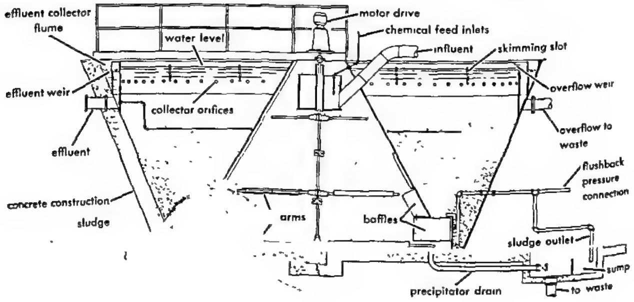


Fig 2 Hot lime-soda water softener, sludge-blanket type.

Hot lime-soda process. The equipment used in this process consists of one or more chemical feeders, a softener unit, and filters. In addition, an integral or separate deaerator may be employed Chemicals used are lime and soda ash. Heating is accomplished with steam (usually at 5–10 psig), and liberated gases are vented to the atmosphere Precipitates produced are calcium carbonate and magnesium hydroxide. The latter reduces the sil-



type with vertical concrete or steel precipitator.

ica content, and if insufficient in quantity, dolomitic lime or activated magnesia may be used instead of lime. A slight excess of soda ash (about 30 ppm) is usually employed, and with this, the total hardness is reduced to less than 25 ppm.

Hot lime-soda and phosphate process. In this two-stage process, the settled water from a hot lime-soda softener is treated with sodium phosphate (either in an integral compartment or in a separate settling tank), which precipitates the residual hardness as calcium and magnesium phosphates. The settled water from this second stage is then filtered. Total hardness of final effluent is "zero" (1-2 ppm).

Hot lime-soda and zeolite process. In this two-stage process, the hard water may be treated with lime or with lime and soda ash in the first stage. In either case, the filtered effluent from the hot lime-soda softener is then passed through a sodium cation exchanger which usually reduces the residual hardness to "zero" (1-2 ppm).

Zeolite process. The equipment used in this process consists of one or more sodium cation-exchanger (zeolite) water softener units and, in most industrial water softeners, a brine-measuring tank and a wet salt storage tank or basin. Hard waters are softened by passing them, usually downwardly, through a columnar bed of a granular or bead-type sodium cation exchanger which removes the calcium and magnesium cations from the water by exchanging for them an equivalent amount of sodium cations. At the end of the softening run, the softener unit is cut out of service and the bed is (1) backwashed; (2) regenerated with a solution of common salt (NaCl), which removes the calcium and magnesium cations by exchanging sodium cations for them, thus restoring the bed to its original sodium state; and (3) rinsed free from calcium and magnesium chlorides and excess salt, after which the unit is returned to service. With most hard waters, the hardness is reduced to practically "zero" (1-2 ppm). Waters very high in hardness or sodium salts are not so completely softened.

Hydrogen cation-exchanger process. The equipment consists of one or more hydrogen cation-exchanger units [illegible] in [illegible] tan[illegible] process except that calcium, magnesium, and sodium cations are removed by exchanging hydrogen cations for them, regeneration is effected with a dilute acid, usually sulfuric; and the effluent contains carbon dioxide and a mineral acid content equivalent to that of the sulfate and chloride ions of the hard water. Most of the carbon dioxide may be removed by aeration or degasification, and the acids may be neutralized with the effluent from a sodium cation exchanger, or neutralized with caustic soda, or removed by an anion exchanger. Hardness removal is practically to "zero." See ION EXCHANGE; SALINE WATER RECLAMATION; SURFACE-ACTIVE AGENT. [E.NO.]

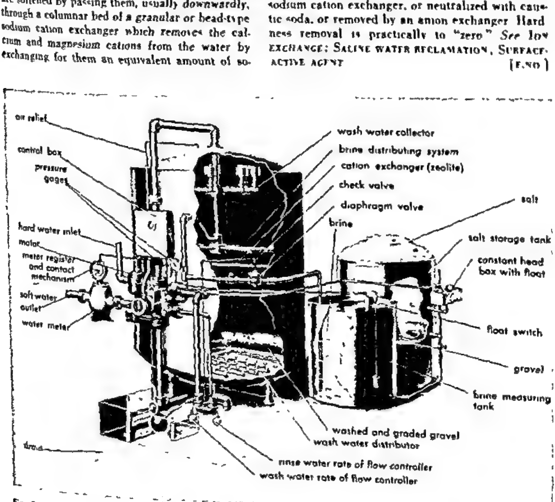


Fig. 3 Sodium cation-exchanger (zeolite) water softener, automatic type

men sometimes believe they are detrimental to sport fishing They produce large litters of living young. They are active either by day or night. *See* REPTILIA. [J.D B]

Water softening

A water-treatment process by which undesirable cations (of calcium and magnesium) are removed from hard waters. The presence of these cations in water is undesirable for household purposes, boiler feed, food processing, and chemical processing, because of reactions that form soap scum, boiler scale, and unwanted by-products. *See* WATER TREATMENT.

Hard waters may be softened by precipitation processes, cation-exchange processes, or combinations of these. The choice of a process depends on a number of factors, among which are the composition of the hard water, the end uses of the softened water, the type or types of hardnesses to be removed, the degree of removal required, and the relative processing costs. In the following discussion, hardness is expressed as calcium carbonate equivalents in parts per million (ppm), and lime means hydrated lime.

Cold lime-soda processes. The equipment used in these processes consists of chemical feeders, a softener unit, and usually filters. Chemicals used are lime (or lime and soda ash) plus a coagulant. Precipitates produced are calcium carbonate and magnesium hydroxide. Lime may be used to reduce the calcium carbonate hardness to about 35 ppm or to reduce both calcium and magnesium carbonate hardness to about 70 ppm. Lime and soda ash will reduce total hardness (carbonate and noncarbonate) to about 70 ppm without using excess chemicals. With excess chemicals, total hardness may be reduced to 16 ppm

The cold lime-soda process may be combined with the zeolite process in a two-stage operation in which the filtered effluent from a cold lime-soda softener is passed through a sodium cation-exchanger (zeolite) water softener. Either lime or lime and soda ash may be used in the first stage In either case, the residual hardness from most hard waters is reduced to practically "zero" hardness (1–2 ppm) by the second stage.

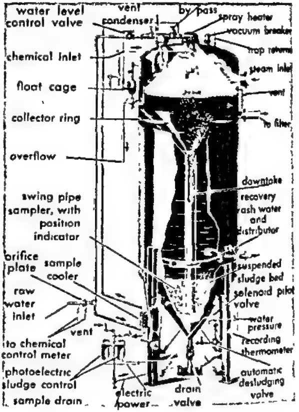
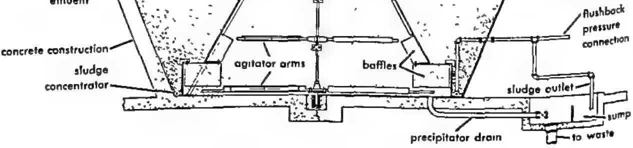


Fig 2. Hot lime-soda water softener, sludge-blanket type.

Hot lime-soda process. The equipment used in this process consists of one or more chemical feeders, a softener unit, and filters. In addition, an integral or separate deaerator may be employed Chemicals used are lime and soda ash Heating is accomplished with steam (usually at 5–10 psig), and liberated gases are vented to the atmosphere Precipitates produced are calcium carbonate and magnesium hydroxide. The latter reduces the sil-



ica content, and if insufficient in quantity, dolomitic lime or activated magnesia may be used instead of lime. A slight excess of soda ash (about 30 ppm) is usually employed, and with this, the total hardness is reduced to less than 25 ppm

Hot lime soda and phosphate process. In this two-stage process, the settled water from a hot lime-soda softener is treated with sodium phosphate (either in an integral compartment or in a separate settling tank), which precipitates the re- [illegible]

"zero" (1–2 ppm).

Hot lime-soda and zeolite process. In this two-stage process, the hard water may be treated with lime or with lime and soda ash in the first stage. In either case, the filtered effluent from the hot lime soda softener is then passed through a sodium cation exchanger which usually reduces the residual hardness to "zero" (1–2 ppm).

Zeolite process. The equipment used in this process consists of one or more sodium cation-exchanger (zeolite) water softener units and, in most industrial water softeners, a brine-measuring tank and a wet salt storage tank or basin. Hard waters are softened by passing them, usually downwardly, through a columnar bed of a granular or bead type sodium cation exchanger which removes the calcium and magnesium cations from the water by exchanging for them an equivalent amount of sodium cations. At the end of the softening run, the softener unit is cut out of service and the bed is (1) backwashed; (2) regenerated with a solution of common salt (NaCl), which removes the calcium and magnesium cations by exchanging sodium cations for them, thus restoring the bed to its original sodium state; and (3) rinsed free from calcium and magnesium chlorides and excess salt, after which the unit is returned to service. With most hard waters, the hardness is reduced to practically "zero" (1–2 ppm). Waters very high in hardness or sodium salts are not so completely softened.

Hydrogen cation-exchanger process. The equipment consists of one or more hydrogen cation- [illegible]

process except that calcium, magnesium, and sodium cations are removed by exchanging hydrogen cations for them, regeneration is effected with a dilute acid, usually sulfuric; and the effluent contains carbon dioxide and a mineral acid content equivalent to that of the sulfate and chloride ions of the hard water. Most of the carbon dioxide may be removed by aeration or degasification, and the acids may be neutralized with the effluent from a sodium cation exchanger, or neutralized with caustic soda, or removed by an anion exchanger. Hardness removal is practically to "zero." *See* ION EXCHANGE; SALINE WATER RECLAMATION; SURFACE-ACTIVE AGENT [E.NO.]

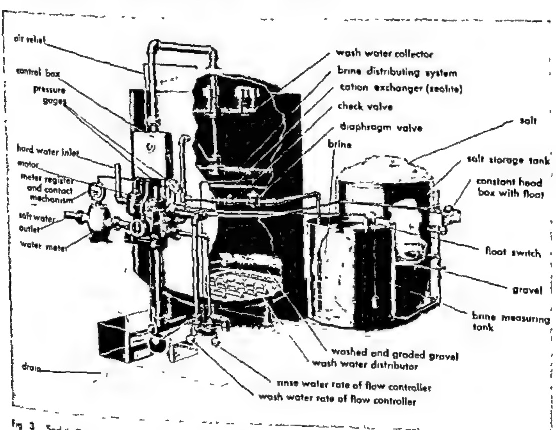
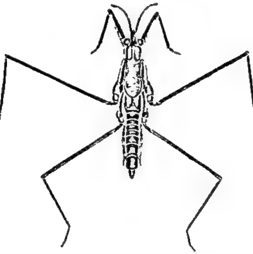


Fig 3 Sodium-[illegible] xchanger (zeolite) water softener, automatic type

Water strider

Any of several species of insects of the family Gerridae, order Hemiptera. Except for one genus they are exclusively fresh-water forms. There are about 20 species in the United States. They are small, slender insects with the middle and hind pairs of legs long and fringed with hairs, enabling them to support themselves on the surface film of the water and to move quickly, as though skating. The front pair of legs is short and modified for grasping food. Water striders feed upon both living and dead insects There are winged and wingless individuals in the same species



The water strider, *Gerris* sp., length to ½ in (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

The genus *Halobates* lives on the ocean surface, sometimes many miles from land, and is the only truly marine insect genus

The water scorpions of the family Nepidae are similar in appearance and habits to the striders, but are larger and distinguished by the long respiratory tube at the end of the abdomen *See* HEMIPTERA. [J.D B]

Water supply engineering

A branch of civil engineering concerned with the development of sources of supply, transmission, distribution and treatment of water. The term is used most frequently in regard to municipal water works, but applies also to water systems for industry, irrigation, and other purposes *See* CIVIL ENGINEERING.

SOURCES OF WATER SUPPLY

Underground waters, rivers, lakes, and reservoirs, the primary sources of fresh water, are replenished by rainfall. Some of this water flows to the sea through surface and underground channels, some is taken up by vegetation, and some is lost by evaporation.

Ground water. Water obtained from subsurface sources, such as sands and gravels and porous or fractured rocks, is called ground water. Ground water flows toward points of discharge in river valleys and, in some areas, along the seacoast. The flow takes place in water-bearing strata known as aquifers. The velocity may be a few feet to several miles per year, depending upon the permeability of the aquifer and the hydraulic gradient or slope. A steep gradient or slope indicates relatively high pressure, or head, forcing the water through the aquifer. When the gradient is flat, the pressure forcing the water is small. Where the velocity is extremely low, the water is likely to be highly mineralized; if there is no movement, the water is rarely fit for use. *See* AQUIFER; GROUND WATER

Permeability is a measure of the ease with which water flows through an aquifer. Coarse sands and gravels, and limestone with large solution passages, have high permeability. Fine sand, clay, silt, and dense rocks (unless badly fractured) have low permeability.

Water table. In an unconfined stratum the water table is the top or surface of the ground water It may be within a few inches of the ground surface or hundreds of feet below. Normally it follows the topography. Aquifers confined between impervious strata may carry water under pressure. If a well is sunk into such an aquifer and the pressure is sufficient, water may be forced to the surface, resulting in an artesian well. The water table elevation and artesian pressure may vary substantially with the seasons, depending upon the amount of rainfall recharging the aquifer and the amount of water taken from the aquifer If pumpage exceeds recharge for an extended period, the aquifer is depleted and the water supply lost.

Salt-water intrusion Normally the ground water flow is toward the sea. This normal flow may be reversed, however, by overpumping and lowering of the water table or artesian pressure in an aquifer Salt water flowing into the fresh-water aquifer being pumped is called salt-water intrusion.

Springs. Springs occur at the base of sloping ground or in depressions where the surface elevation is below the water table, or below the hydraulic gradient in an artesian aquifer from which the water can escape Artesian springs are fed through cracks in the overburden or other natural channels extending from the confined aquifer under pressure to the surface *See* SPRING (WATER).

Wells. Wells are vertical openings, excavated or drilled, from the ground surface to a water-bearing stratum or aquifer. Pumping a well lowers the water level in it, which in turn forces water to flow from the aquifer. Thick, permeable aquifers may yield several million gallons daily with a drawdown (lowering) of only a few feet Thin aquifers, or impermeable aquifers, may require several times as much drawdown for the same yield, and frequently will yield only small supplies

Dug wells, several feet in diameter, are frequently used to reach shallow aquifers, particularly for small domestic and farm supplies. They furnish small quantities of water, even if the soils penetrated are relatively imp Large-capacity

dug wells or caisson wells, in coarse sand and gravel, are used frequently for municipal supplies. Drilled wells are sometimes several thousand feet deep

The portion of a well above the aquifer is lined with concrete, stone, or steel casing, except where the well is through rock that stands without support. The portion of the well in the aquifer is built with open joint masonry or screens to admit the water into the well Metal screens, made of perforated sheets or of wire wound around supporting ribs, are used most frequently The screens are galvanized iron, bronze, or stainless steel, depending upon the corrosiveness of the water and the expected life of the well *See* WELL.

The distance between wells must be sufficient to avoid harmful interference when the wells are pumped In general, economical well spacing varies directly with the quantity of water to be pumped, and inversely with the permeability and thickness of the aquifer. It may range from a few feet to a mile or more.

Infiltration galleries are shafts or passages extending horizontally through an aquifer to intercept the ground water They are equivalent to a row of closely spaced wells and are most successful in thin aquifers along the shore of rivers, at depths of less than 75 ft The galleries are built in open cuts or by tunneling, usually with perforated or porous liners to screen out the aquifer material and to support the overburden

Ranney wells consist of a center caisson with horizontal, perforated pipes extending radially into the aquifer They are particularly applicable to the development of thin aquifers at shallow depths

Specially designed pumps, of small diameter to fit inside well casings, are used in all well installations, except in flowing artesian wells or where the water level in the well is high enough for direct suction lift by a pump on the surface (about 15 ft) Well pumps are set some distance below the water level, so that they are submerged even after the drawdown is established Well-pump settings of 100 ft are common, and they may exceed 300 ft where the ground-water level is low Multiple-stage centrifugal pumps are used most generally They are driven by motors at the surface through vertical shafts, or by waterproof motors attached directly below the pumps Wells are sometimes pumped by air lift, that is by injecting compressed air through a pipe to the bottom of the well

Surface water. Natural sources, such as rivers and lakes, and impounding reservoirs are sources of surface water See RESERVOIR, SURFACE WATER

Water is withdrawn from rivers lakes and reservoirs through intakes The simplest intakes are pipes extending from the shore into deep water, with or without a simple crib and screen over the outer end Intakes for large municipal supplies may consist of large conduits or tunnels extending to elaborate cribs of wood or masonry containing screens gates and operating mechanisms Intakes in reservoirs are frequently built as integral parts of the dam and may have multiple ports at several levels to permit selection of the best water. The location of intakes in rivers and lakes must take into consideration water quality, depth of water, likelihood of freezing and possible interference with navigation Reservoir intakes are usually designed for gravity flow through the dam or its abutments, In lakes and rivers, the water flows by gravity through the intake to a pumping station on the shore.

TRANSMISSION AND DISTRIBUTION

The water from the source must be transmitted to the community or area to be served and distributed to the individual customers.

Transmission mains. The major supply conduits, or feeders, from the source to the distribution system are called mains or aqueducts.

Canals The oldest and simplest type of aqueducts, especially for transmitting large quantities of water, are canals Canals are used where they can be built economically to follow the hydraulic gradient or slope of the flowing water. If the soil is suitable, the canals are excavated with sloping sides and are not lined Otherwise, concrete or asphalt linings are used Gravity canals are carried across streams or other low places by wooden or steel flumes, or under the streams by pressure pipes known as [illegible] See CANAL

[illegible]

tunnels may be lined to prevent the overburden from collapsing to prevent leakage, or to reduce friction losses by providing a smooth interior. See TUNNEL.

Pipelines Pipelines are a common type of transmission main, especially for moderate supplies not requiring large aqueducts or canals Pipes are of cast iron, steel, reinforced concrete, cement asbestos, or wood Pipeline material is determined by cost, durability, ease of installation and maintenance, and resistance to corrosion The pipeline must be large enough to deliver the required amount of water and strong enough to withstand the maximum gravity or pumping pressure Pipelines are usually buried in the ground for protection and coolness *See* PIPELINE.

Distribution system. Included in the distribution system are the network of smaller mains branching off from the transmission mains, the house services and meters, the fire hydrants, and the distribution storage reservoirs The network is composed of transmission or feeder mains, usually 12 in or more in diameter, and lateral mains along each street, or in some cities along alleys between the streets The mains are installed in grids so that lateral mains are fed from both ends where possible Mains fed from one direction only are called dead ends, they are less reliable and will not furnish as much water for fire protection as mains within the grid. Valves at intersections of mains permit a leaking or damaged section of pipe

to be shut off with minimum interruption of water service to adjacent areas.

House services. The small pipes, usually of iron, copper, or plastic material, extending from the water main in the street to the customer's meter at the curb line or in the cellar are called house services. In most cities each service is metered, and the customer's bill is based on the water actually used.

Fire hydrants. Fire hydrants have a vertical barrel extending to the depth of the water main, a quick-opening valve with operating nut at the top, and connections threaded to receive fire hose. Hydrants must be reliable, and they must drain upon closing to prevent freezing.

Distribution reservoirs. These are used to supplement the source of supply and transmission system during peak demands, and to provide water during a temporary failure of the supply system. In small water works the reservoirs usually equal at least one day's water consumption; in larger systems the reservoirs are relatively smaller but adequate to meet fire-fighting demands. Ground storage reservoirs, elevated tanks, and standpipes are used for distribution reservoirs.

Ground storage reservoirs, if on high ground, can feed the distribution system by gravity, but otherwise it is necessary to pump water from the reservoir into the distribution system. Circular steel tanks and basins built of earth embankments, concrete, or rock masonry are used. Earth reservoirs are usually lined to prevent leakage and entrance of dirty water. The reservoirs should be covered to protect the water from dust, rubbish, and bird droppings, but many older reservoirs without covers are in use

Elevated storage reservoirs are tanks on towers, or high cylindrical standpipes resting on the ground. Storage reservoirs are built high enough so that the reservoir will maintain adequate pressure in the distribution system at all times

Elevated tanks are usually of steel plate, mounted on steel towers Wood is sometimes used for industrial and temporary installations Standpipes are made of steel plate, strong enough to withstand the pressure of the column of water. The capacity of standpipe required is greater than an elevated tank because only the upper portion of a standpipe is at high enough elevation for normal use.

Distribution-system design. To assure the proper location and size of feeder mains and laterals to meet normal and peak water demands, a distribution system must be expertly designed. As the water flows from the source of supply or distribution reservoir across a city, the water pressure is lowered by the friction in the pipes. The pressures required for adequate service depend upon the height of buildings, need for fire protection, and other factors, but 40–60 psi is the minimum for good service. Higher pressures for fire fighting are obtained by booster pumps on fire engines which take water from fire hydrants. In small towns adequate hydrant flows are the controlling factor in determining water-main size; in larger communities the peak demands for air conditioning and lawn sprinkling during the summer months control the size of main needed The capacity of a distribution system is usually determined by opening fire hydrants and measuring simultaneously the discharge and the pressure drop in the system The performance of the system when delivering more or less water than during the test can be computed from the pressure drops recorded.

An important factor in the economical operation of municipal water supplies is the quantity of water lost from distribution because of leaky joints, cracked water mains, and services abandoned but not properly shut off. Unaccounted-for water, including unavoidable slippage of customers' meters, may range from 10% in extremely well-managed systems to 30–40% in poor systems. The quantities flowing in feeder mains, friction losses, and the amount of leakage are frequently measured by means of pitometer surveys. A pitometer is a portable meter that can be inserted in a water main under pressure to measure the velocity of flow, and thus the quantity of flow. *See* FLOW MEASUREMENT

Pumping stations. Pumps are required wherever the source of supply is not high enough to provide gravity flow and adequate pressure in the distribution system. The pumps may be high or low head depending upon the topography and pressures required. Booster pumps are installed on pipelines to increase the pressure and discharge, and adjacent to ground storage tanks for pumping water into distribution systems. Pumping stations usually include two or more pumps, each of sufficient capacity to meet demand when one unit is down for repairs or maintenance. The station must also include piping and valves arranged so that a break can be isolated quickly without cutting the whole station out of service

Centrifugal pumps have displaced steam-driven reciprocating pumps in modern practice, although many of the old units continue to give good service The centrifugal pumps are driven by electric motors, steam turbines, or diesel engines, with gasoline engines frequently used for standby service The centrifugal pumps used most commonly are designed so that the quantity of water delivered decreases as the pumping head or lift increases. Both horizontal and vertical centrifugal pumps are available in a wide capacity range. In the first type, the pump shaft is horizontal with the driving motor or engine at one end of the pump. Vertical pumps are driven by a vertical-shaft motor directly above the pump, or by a horizontal engine through a right-angle gear head *See* PUMP

Automatic control of pumping stations is provided to adjust pump operations to variations in water demand The controls start and stop pumps of different capacity as required. In the event of mishap or failure of a unit, alarms are sounded The controls are activated by the water level in a reservoir or tank, by the pressure in a water main, or by the rate of flow through a meter. Remote control of pumps is frequently [illegible]yed, with the sig-

nals transmitted over telephone wires. See WATER TREATMENT. [R H]

Bibliography: R. W. Abbett, *American Civil Engineering Practice*, 3 vols., 1956; C. V. Davis, *Handbook of Applied Hydraulics*, 2d ed., 1952; G. M. Fair and J. L. Geyer, *Water Supply and Waste Disposal*, 1954

Water table

The upper surface of the zone of saturation in permeable rocks not confined by impermeable rocks. It may also be defined as the surface underground at which the water is at atmospheric pressure. Saturated rock may extend a little above this level, but the water in it is held up above the water table by capillarity and is under less than atmospheric pressure; therefore, it is the lower part of the capillary fringe and is not free to flow into a well by gravity. Below the water table the water is free to move under the influence of gravity. The position of the water table is shown by the level at which water stands in wells penetrating an unconfined water-bearing formation.

Where a well penetrates only impermeable material, there is no water table and the well is dry. But if the well passes through impermeable rock into water-bearing material whose hydrostatic head is higher than the level of the bottom of the impermeable rock, water will rise approximately to the level it would have assumed if the whole column of rock penetrated had been permeable. This is called artesian water, and the surface to which it rises is called the piezometric surface. See ARTESIAN SYSTEMS.

The water table is not a level surface but has irregularities that are commonly related to, though less pronounced than, those of the land surface. Also, it is not stationary but fluctuates with the seasons and from year to year. It generally declines during the summer months when the vegetation uses most of the water that falls as precipitation, and rises during the late winter and spring when the demands of vegetation are low. It usually reaches its lowest point after the end of the growing season and its highest point just before the beginning of the growing season. Superimposed on the annual fluctuations are fluctuations of longer period which are controlled by climatic variations. The water table is also affected by withdrawals, as by pumping from wells. See GROUND WATER.

[A N S]

Water treatment

Physical and chemical processes for making water suitable for human consumption and other purposes. Drinking water must be bacteriologically safe and comparatively free of turbidity, color, and taste-producing substances. Excessive hardness and high concentration of dissolved solids are also undesirable, particularly for boiler feed and industrial purposes. The more important treatment processes are sedimentation, coagulation, filtration, disinfection, softening, and aeration.

Plain sedimentation. Silt, clay, and other fine material settle to the bottom if the water is allowed to stand or flow quietly at low velocities. Sedimentation occurs naturally in reservoirs and is accomplished in treatment plants by basins or settling tanks. The detention time in a settling basin may range from an hour to several days. The water may flow horizontally through the basin, with solids settling to the bottom, or may flow vertically upward at a low velocity so that the particles will settle through the rising water. Settling basins are most effective if shallow, and rarely exceed 10–20 ft in depth. Plain sedimentation will not remove extremely fine or colloidal material within a reasonable time, and the process is used principally as a preliminary to other treatment methods.

Coagulation. Fine particles and colloidal material are combined into masses by coagulation. These masses, called floc, are large enough to settle in basins and to be caught on the surface of filters. Waters high in organic material and iron may coagulate naturally with gentle mixing. The term is usually applied to chemical coagulation, in which iron or aluminum salts are added to the water to form insoluble hydroxide floc. The floc is a feathery, highly absorbent substance to which color-producing colloids, bacteria, fine particles and other substances become attached and are thus removed from the water.

The coagulant dose is a function of the physical and chemical character of the raw water, the adequacy of settling basins and filters, and the degree of purification required. Moderately turbid water coagulates more easily than perfectly clear water, but extremely turbid water requires more coagulant. Coagulation is more effective at higher temperatures. Lime, soda ash, or caustic soda may be required in addition to the coagulant to provide sufficient alkalinity for the formation of floc, and regulation of the pH (hydrogen ion concentration) is usually desirable for best results. Powdered limestone, clay, bentonite, or silica are sometimes added as coagulant aids to strengthen and weight the floc. Coagulation and sedimentation are the most important parts of modern water purification.

Filtration. Suspended solids, colloidal material, bacteria, and other organisms are filtered out by passing the water through a bed of sand or pulverized coal, or through a matrix of fibrous material supported on a perforated core. Filtration of turbid ...

... sand filters. Known also as English filters, these consist of beds of sand 20–48 in. deep, through which the water is passed at fairly low rates—2.5–10 × 10^6 gallons per acre. The size of beds ranges from a fraction of an acre in small plants to several acres in large plants. An underdrain system of graded gravel and perforated pipes transmits the filtered water from the filters to the point of discharge. The sand is usually fine, rang-

ing from 0.2–0.5 mm in diameter. The top of the filter clogs with use, and a thin layer of dirty sand is scraped from the filter periodically to maintain capacity.

Slow sand filters operate satisfactorily with reasonably clear waters but clog rapidly with turbid waters. The filters are covered in cold climates to prevent the formation of ice and to facilitate operation in the winter. In milder climates they are often open. Slow sand filters have a high bacteriological efficiency, but few have been built since the development of water disinfection, because of the large area required, the high construction cost, and the labor needed to clean the filters and to handle the filter sand.

Rapid sand filters. These operate at rates of 125–250 × 10^6 gallons per acre per day; or 25 to 50 times slow sand filter rates. The high rate of operation is made possible by coagulation and sedimentation ahead of filtration to remove the heaviest part of the load, the use of fairly coarse sand, and facilities for backwashing the filter to keep the bed clean. The filter beds are small, generally ranging from 150 square feet (ft^2) in small plants to 1,500 ft^2 in the largest filter plants. The filters consist of a layer of sand or occasionally crushed anthracite coal 18–24 in. deep, resting on graded layers of gravel above an underdrain system. The sand is coarse, 0.4–1.0 mm in diameter, depending upon the raw water quality and pretreatment, but the grain size must be fairly uniform to assure proper backwashing. The underdrain system serves both to collect the filtered water and to distribute the wash water under the filters when they are being washed. Several types of underdrains are used, including perforated pipes, perforated false bottoms of concrete and tile and porous plates

Filters are backwashed at rates 5–10 times the filtering rate. The wash water passes upward through the sand and out of the filters by way of wash-water gutters and drains Washing agitates the sand bed and releases the dirt to flow out of the filter with the wash water The quantity of water used for washing ranges from 1 to 10% of the total output, depending upon the turbidity of the water applied to the filters and the efficiency of filter design.

Municipal and large-capacity filters for industry usually are built in concrete boxes or in open tanks of wood and steel. The flow through the sand may be caused by gravity, or the water may be forced through the sand under pressure by pumping. Pressure filters can be operated at higher rates than gravity filters, because of the greater head available to force the water through the sand. However, excessive pressure causes turbidity, and bacteria may appear in the discharge water. For this reason, and because pressure filters are difficult to inspect and keep in good order, open gravity filters are favored for public water supplies

Diatomaceous earth filters. Swimming pool installations and small water supplies frequently use this type of filter. The filters consist of a medium or septum supporting a layer of diatomaceous earth through which the water is passed. A filter layer is built up by the addition of diatomaceous earth to the water. When the pressure loss becomes excessive, filters must be backwashed and a fresh layer of diatomaceous earth applied Filter rates of 2½–6 gallons per minute per square foot are attained.

Disinfection. There are several methods of treatment of water to kill living organisms, particularly pathogenic bacteria; the application of chlorine or chlorine compounds is the most common Less frequently used methods include the use of ultraviolet light, ozone, or silver ions. Boiling is the favorite household emergency measure.

Chlorination is simple and inexpensive and is practiced almost universally in public water supplies. It is often the sole treatment of clear, uncontaminated waters In most water treatment plants it supplements coagulation and filtration. Chlorination is used also as protection against contamination of water in distribution mains and reservoirs after purification

Chlorine gas is most economical and easiest to apply in large systems. For small works, calcium hypochlorite or sodium hypochlorite is frequently used. Regardless of which form is used, the dose varies with the water quality and degree of contamination Clear, uncontaminated water can be disinfected with small doses, usually less than one part per million; contaminated water may require several times as much. The amount of chlorine taken up by organic matter and minerals in water is known as the chlorine demand. For proper disinfection the dose must exceed the demand so that free chlorine remains in the water.

Chlorination alone is not reliable for the treatment of contaminated or turbid water. A sudden increase in the chlorine demand may absorb the full dose and provide no residual chlorine for disinfection. and it cannot be assumed that the chlorine will penetrate particles of organic matter. Chlorine is applied before filtration, after filtration, and sometimes at both places.

Chlorine sometimes causes objectionable tastes or odors in water. This may be due to excessive chlorine doses, but more frequently it is caused by a combination of chlorine and organic matter, such as algae, in the water Some algae, relatively unobjectionable in the natural state, produce unbearable tastes after chlorination In other cases, strong chlorine doses oxidize the organic matter completely and produce odor-free water. Excessive chlorine may be removed by dechlorination with sulfur dioxide Also, ammonia is often added for taste control to reduce the concentration of free chlorine Activated carbon is also effective in the reduction of both natural and chlorine tastes and odors

Water softening. The "hardness" of water is due to the presence of calcium and magnesium salts. These salts make washing difficult, waste soap, and cause unpleasant scums and stains in house-

holds and laundries. They are especially harmful in boiler feed water because of their tendency to form scales

Municipal water softening is common where the natural water has a hardness in excess of 150 parts per million. Two methods are used. (1) the water is treated with lime and soda ash to precipitate the calcium and magnesium as carbonate and hydroxide, after which the water is filtered, (2) the water is passed through a porous cation exchanger which has the ability of substituting sodium ions in the exchange medium for calcium and magnesium in the water. The exchange medium may be a natural sand known as [illegible]

For high pressure steam boilers or some other industrial processes, almost complete deionization of water is needed, and treatment includes both cation and anion exchangers. Lime soda plants are similar to water *purification* plants, with coagulation, settling and filtration. Zeolite or cation-exchange plants are usually built of steel tanks with appurtenances for backwashing the media with salt brine. If the water to be softened is turbid, filtration ahead of zeolite softening may be required. *See* WATER SOFTENING.

Aeration. Aeration is a process of exposing water to air by dividing the water into small drops, by forcing air through the water, or by a combination of both. The first method uses jets, fountains, waterfalls and riffles; in the second, compressed air is admitted to the bottom of a tank through perforated pipes or porous plates, in the third, drops of water are met by a stream of air produced by a fan [illegible]

re [illegible]

pr [illegible]

ir [illegible] the iron ahead of sedimentation or filtration. *See* WATER SUPPLY ENGINEERING

[R K]

Bibliography. R. W Abbett, *American Civil Engineering Practice*, 3 vols., 1956; C. V Davis (ed), *Handbook of Applied Hydraulics*, 2d ed , 1952, G. M. Fair and J. C. Geyer, *Water Supply and Waste Water Disposal*, 1954

Water tunnel

A hydrodynamic test and research tool or facility, comprising a well guided and controlled stream of water. The water tunnel is related to towing tanks but differs from them by the fact that in the towing tank the test object moves and the water is essentially at rest, whereas in the water tunnel the test object is at rest (or rotating, as for example, a propeller) and the water is moving (*see* TOWING TANK). With respect to arrangement and operation, the water tunnel is similar to the subsonic wind tunnel, except for the difference in the test medium (*see* WIND TUNNEL). There are few water tunnels as compared with the number of wind tunnels in operation.

Construction. A typical water tunnel is shown in Fig 1. This type forms a closed loop of conduit with a carefully designed, strongly contracted test, or working, section, a circulating pump, and corner elbows shaped as they are in wind tunnels to minimize flow distortions and disturbances. The contraction in front of the test section further serves to produce a stream with a uniform velocity distribution. The size of a water tunnel is usually expressed as the diameter of its test section. Existing water tunnels range in size from a few inches to 48 in. The test section is on the upper leg to facilitate lowering the water level below it for insertion of test objects (Fig 2).

An essential characteristic of most water tunnels is the possibility of changing the absolute pressure in the tunnel without necessarily changing the velocity of the test stream. This characteristic is important for investigations involving cavitation

Most water tunnels are operated with the entire test circuit filled with water. Consequently they can serve only the investigations of completely submerged bodies or of conduits and machinery such as pumps or turbines that are completely filled with water. A few tunnels have been built to operate with a free water surface in the test section. Such free-surface water tunnels are used to investigate open-surface structures such as ship models, with the model standing still. Free-surface tunnels can be used for cavitation studies only if the air space above the surface is enclosed to permit changes in absolute pressure

Many water tunnels are equipped with special means for the elimination of undissolved gas liberated by cavitation. The most common device used for this purpose is the reabsorber, introduced in the late 1940s by R. T Knapp. It consists of a downward extension of the tunnel circuit, which subjects the tunnel stream for a fair length of time to an increased pressure in order to absorb undissolved air into the water. This device therefore leaves the total gas content of the tunnel water unchanged.

Applications. Principal applications of water tunnels are the investigation of submerged bodies, of hydrodynamic machinery, and of basic flow phenomena in liquids

With respect to submerged bodies, the water tunnel serves the same purpose as the wind tunnel with respect to airplane models. The measurement of forces acting on the test body in the test stream requires water tunnel balances similar to wind tunnel balances. The obvious requirement to keep the water tunnel completely sealed and to prevent effects of changes in absolute tunnel pressure on the balance readings makes the construction of water tunnel balances more difficult than that of balances for subsonic wind tunnels. Water tunnel balances arranged internally as well as externally to the test body are in successful use

Water tunnels for the investigation of hydrodynamic machinery may differ in shape from the

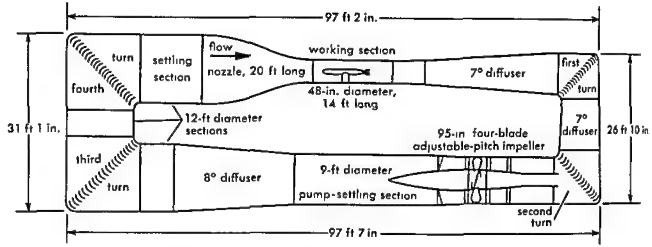


Fig 1. Elevation of world's largest water tunnel.

previously described arrangement, with the test machine taking the place of the straight test section. Such water tunnels are often called test stands, specifically cavitation test stands Water tunnels with straight test sections, as described before, are used regularly for the investigation of marine propellers

Water tunnels, like wind tunnels, are essential tools for basic hydrodynamic investigations The phenomena that can be investigated in water tunnels are the same as those that can be investigated in subsonic wind tunnels, such as force action between the stream and the test body, boundary layer phenomena, and turbulence. In addition, water tunnels permit investigation of phenomena peculiar to hydrodynamics, particularly cavitation, hydroelastic phenomena, and free-surface phenomena.

Cavitation. The vaporization of the water due to local pressure reductions under the dynamic action of the flow is cavitation. It depends, therefore, upon the absolute pressure in the test stream (*see* CAVITATION) This phenomenon sets a limit for the velocities of flow obtainable under certain conditions and is therefore of great practical importance (*see* HYDRODYNAMICS) It is observed in water tunnels optically, as well as acoustically, by use of the capability of water tunnels to change the absolute pressure of the test stream without changing other flow characteristics Cavitation is usually detectable as a white foamlike region Furthermore, sensitive, high-frequency hydrophones are an excellent detector of cavitation in water tunnels

Hydroelasticity. The interaction between the elastic behavior of the test body and the flow of the water is hydroelasticity. It is similar to aeroelasticity, but the flow phenomena that dominate hydroelasticity are different from those dominating aeroelasticity, because of the great difference in the densities of the flowing media Hydroelasticity is one of the most important water tunnel investigations.

Free-surface phenomena. Free-surface water tunnels permit the investigation of free-surface phenomena. They should be used in preference to towing tanks only where cavitation is important or where it is desirable to hold test object stationary

Instruments and limitations. The principal tools of water tunnel investigations include various types of pressure gages, water tunnel balances, torque and thrust meters for rotating machinery, and instruments for acoustic observations of cavitation, hydroelastic vibration, and turbulence. In 1959 the Schlieren method was used for the first time successfully for the observation of certain flow phenomena in a water tunnel. *See* SCHLIEREN PHOTOGRAPHY.

The limitations of water tunnels are essentially the same as those of wind tunnels: limitations in the size of the test object due to interference of the flow field with the boundaries of the test stream, and limitations regarding the uniformity of the test stream by tunnel-generated turbulence. Various means have been tried to reduce tunnel interference For bodies of revolution a change in the shape of the test section so as to simulate a stream

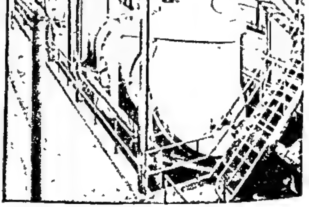


Fig 2 Water tunnel at Pennsylvania State University; dimensions given in Fig. 1.

surface in an infinitely extended medium has been used successfully. The suppression of tunnel turbulence by means of screens is similar to wind tunnel practice. In addition to the limitations that they have in common with wind tunnels, water tunnels are limited by the liberation of gas in connection with extensive cavitation and by acoustic interference between the test stream and the tunnel walls. [C.F.W.]

Bibliography: J. M. Robertson, Water tunnels for hydraulic investigations, *Trans. ASME*, 78(1): 95–104, 1956

Water-borne disease

Disease transmitted by drinking water or by contact with potable or bathing water. Certain bacterial, protozoan, and helminthic diseases of man are commonly or exclusively water-borne. The bacterial and protozoan pathogens are not normal to the water environment and do not multiply in natural waters. They gain entrance to water primarily from pollution by human excrement and less significantly from pollution by domestic and wild animals. In contrast, the helminthic pathogens may have a water animal as an intermediate host and may have a temporary free-living stage in water. See ENTERIC BACILLI; EPIDEMIOLOGY, INFECTIOUS DISEASE TRANSMISSION; PARASITOLOGY, MEDICAL. [S.C.R.]

Water-borne disease

Type	Disease	Agent
Bacterial	Typhoid fever	*Salmonella typhosa*
	Asiatic cholera	*Vibrio comma*
	Bacillary dysentery	*Shigella* species
Protozoan	Amebic dysentery	*Endamoeba histolytica*
Helminthic	Ascariasis (roundworm infection)	*Ascaris lumbricoides*
	Trichuriasis (whipworm infection)	*Trichuris trichura*
	Schistosomiasis	*Schistosoma* species
	Dracontiasis	*Dracunculus medinensis*

Watermelon

The edible fruit of *Citrullis vulgaris*, native to Africa. See FRUIT (BOTANY). Leaves are deeply 5–7 lobed; flowers are light yellow. See FLOWER (BOTANY); LEAF (BOTANY). Among the common American varieties, fruit characters show the following variations: shape, round to cylindrical-oblong; weight, 5–40 lbs or more; rind color, very light to very dark green and in some species striped light and dark green; flesh color, pink to red or yellow (the preserving melon, or citron, is white-fleshed); seed colors, tan to brown, black, red, green, and speckled. The rind is smooth and relatively hard; the interior is completely filled with sweet, fragile juicy tissue containing many seeds. See SEED (BOTANY).

In 1938 Japanese scientists introduced seedless watermelons. These are grown from seeds developed in fruits resulting from fertilization of flowers on tetraploid plants with pollen of diploid plants (see GENETICS). These seeds are triploid and they produce plants bearing sterile flowers. Fruit formation is stimulated by pollination of these flowers, but the fruits are seedless.

Watermelons are grown extensively in about 20 states of the United States, mostly in the South. Florida, Texas, Georgia, South Carolina, and California produce about 80% of the nation's total. The average annual farm value of the crop from 1949 to 1957 was nearly $40,000,000. See MELON GROWING. [V.R.B.]

Waterspout

An intensely whirling, funnel-shaped vortex, extending several hundred to several thousand feet from a cumulus type cloud down to water surface. The visible funnel consists mostly of atmospheric water vapor condensed because of lower pressure in the vortex; water and salt spray drawn from the underlying water surface also contribute to the structure and visibility of the funnel. Diameters range from a few feet to several hundred, with winds occasionally strong enough to overturn small vessels.

Most waterspouts occur in moist tropical air under unstable conditions favorable for thunderstorms, from which they frequently hang. Nearly all dissipate rapidly on passing inland. Although most common in the tropical oceans, waterspouts are observed in higher latitudes, particularly in summer, and they also occur over inland lakes. [C.W.N.]

Waterspout—an oblique aerial view. (Official U.S. Navy photo)

Water-tube boiler

A steam boiler in which water circulates within tubes and heat is applied from outside the tubes. The outstanding feature of this boiler is that the pressure parts exposed for heat absorption are in the form of relatively small tubes connected to a steam separating drum, so located as to be shielded from high-temperature sources, thus minimizing the hazard of explosion. With such protection of the drum, any probable failure of the heated elements is limited in size and rate of energy escape, thus minimizing the hazard of explosion.

Water-tube construction makes possible enlarged capacity of the boiler, by increasing the length and number of tubes, and greater working pressure, since the tubes of small diameter and nominal thickness can withstand high internal pressure, and the protected drum can be made thicker without risk of overheating Moreover, the versatility of arrangement permits better accommodation of furnace, superheater, and other heat recovery components, with consequent improvement in over-all thermal efficiency.

Early water-tube boilers varied considerably in detail, according to the designs of different manufacturers, but have been broadly divided into straight-tube and bent-tube classification Both consisted essentially of banks of parallel tubes, interconnected by headers or drums The several commercial types became well standardized for different capacities, and were usually sold as discrete units for application to separately constructed firebrick furnaces. Some variations were required to accommodate superheaters of different sizes in the boiler setting.

The straight-tube boiler (Fig. 1), sometimes called the header type boiler, has the advantage of direct accessibility for internal inspection and cleaning through hand-holes, located opposite each tube end in the headers. Generating sections are joined to one or more steam-and-water drums, set parallel to or transversely above the tube bank Positive circulation, which assures complete wetting, as well as uniform temperature and expansion of metal parts in the assembly, is produced by the greater density of water in the downcomer connections compared to the steam-and-water mixture in the riser portion of the circuit. Gas flow between tubes of the bank is directed by baffles, to procure higher mass flow and longer travel for improved heat transfer and absorption efficiency.

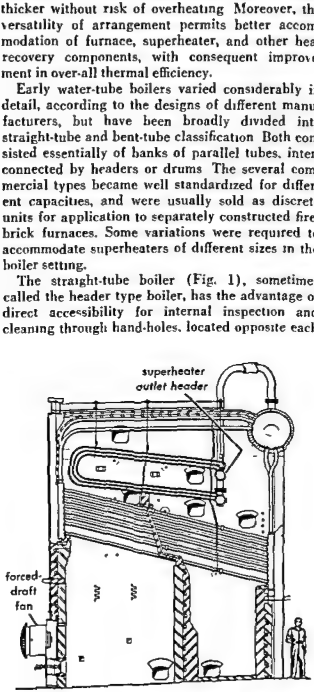


Fig. 1 Straight-tube type of water-tube boiler.

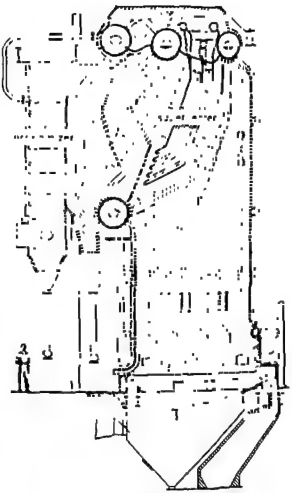


Fig 2 Bent-tube type of water-tube boiler

Bent-tube boilers (Fig. 2), also called drum type boilers, eliminated the multiplicity of small handhole closures, which sometimes leaked, by connecting all tubes into upper and lower drums having relatively few access openings. Although internal inspection of tubes was at first restricted, mechanical cleaners which can traverse the bends are now available, and the development of water treatment and chemical cleaning methods has overcome all the early objections to the use of bent tubes.

Circulation in bent-tube boilers is similarly produced by differential densities of water and steam mixtures in the tubes forming the bank. The predominant down-flow occurs in the regions of cooler gases and is affected by the placement of baffles In recent high-duty bent-tube boilers, active flow is assured in all tubes by selective grouping or by use of shielded or external do[illegible] . See STEAM GENERATING UNIT. [F C

Watt

The unit of power in the meter-kilogram-second (mks) system of units One watt equals 1 joule per second or 10^7 ergs/sec, and 746 watts equals 1 horsepower. In electricity, 1 watt is the power developed in a circuit when the current is 1 ampere and the applied potential difference is 1 volt.

The watt is a convenient unit for many measurements, related units may be used with very small or very large powers For example, the milliwatt and microwatt represent 10^{-3} and 10^{-6} watt, respectively, while 1 kilowatt and 1 megawatt are powers, respectively, of 10^3 and 10^6 watts.

The watt is the unit of power ordinarily employed in mechanics and electricity. In heat measurements, however, calories and British thermal units are often used as energy units, with the corresponding power units of calories (cal) per second and British thermal units (Btu) per second These units are related by the equations

$$1 \text{ watt} = 0.239 \text{ cal/sec}$$
$$1 \text{ watt} = 0.000948 \text{ Btu/sec}$$

See British thermal unit (Btu), Calorie.

Sometimes energy is measured in units which are products of the watt and a time unit. Two such energy units are the watt second and the kilowatt-hour *See* Power

[r w s]

Watt-hour meter

An electricity meter which measures and registers the integral, with respect to time, of the power in the circuit in which it is connected In effect, it is an electric motor the torque of which is proportional to the electric power in the circuit The speed of the rotor is proportional to the torque, making each revolution of the rotor a measurement in watt hours Summation of the watt-hours is accomplished by gearing a counter, or register, to the rotor

The basic elements of a watt-hour meter are the stator, rotor, retarding magnet or magnets, register, and meter housing

Principle of operation. Watt-hour meters may be classified into three types according to fundamental differences in principle of operation

Mercury type watt-hour meter This type is used for measuring energy on a direct-current circuit It differs from other types in that the driven portion of the rotor consists of a radially-slotted copper disk immersed in mercury The load current flows diametrically through the disk interacting with a flux produced by the line-voltage electromagnet to cause the disk to rotate

Mercury type meters are readily applied to high-current loads, since they are used with shunts and only a portion of the load current passes through the current circuit of the meter

Commutator type watt-hour meter This is also used for measuring energy on a direct-current circuit. It may also be used on alternating-current circuits if all windings are of air-core construction This meter is a shunt-type motor The field coils, part of the stator, produce a field proportional to the load current. The armature is mounted on the rotor The armature is energized by the line voltage through a commutator and brushes, producing a rotor torque proportional to the power in the circuit

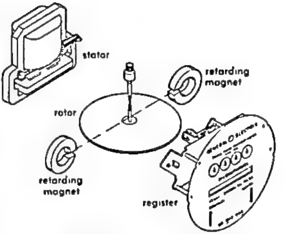


Fig 1 Basic elements of induction type watt-hour meter (*General Electric Co*)

Induction type watt-hour meter This is the common meter found in homes It is used for measuring energy on an ac circuit. Figure 1 is a schematic sketch showing the basic elements of an induction meter Figure 2 shows the stator in more detail.

The potential-circuit winding of the stator is made highly inductive to obtain a quadrature-time relationship between the potential-circuit and the current-circuit working fluxes in the disk air gap. These fluxes, displaced in time and space, produce a rotor-disk torque proportional to the circuit power Retarding magnets control the rotor speed, making it proportional to the power. The register, which is geared to the rotor, records the watt hours.

This type of meter has been developed to a high degree of accuracy under extreme environmental conditions and over great ranges of load and volt-

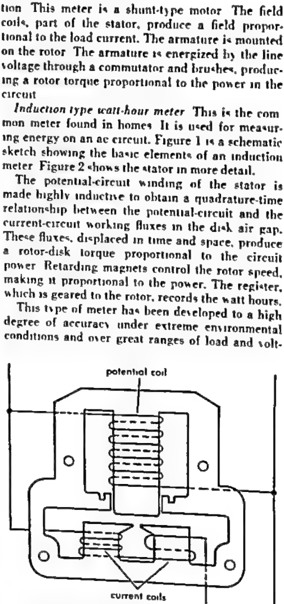


Fig 2 Stator of induction-type watt-hour meter. (*General Electric Co*)

age. While it has the appearance of simplicity and low cost, its magnetic circuitry is extremely complex. Its calibration is stable and maintenance requirements are practically eliminated.

Multistator, or polyphase, watt-hour meters employ the same principle of operation as single-stator meters. Magnetic interference between stators and the need for an adjustment to balance the torques of all stators are added complications of multistator meters.

Special watt-hour meters. Many types of watt-hour meter are available for special needs and applications.

The switchboard-type meter is used for industrial or central-station applications. These differ only in housing construction.

The totalizing meter records in one meter the energy used in two or more circuits. This meter may have four or more stators acting on a single rotor.

The portable watt-hour meter standard is a specially developed, high-accuracy watt-hour meter, having a multiplicity of current and voltage circuits. It is generally used in the meter shops of utilities and in the field for testing watt-hour meters

A combination watt-hour meter and time switch, consisting of a standard single-phase meter and a time switch combined in one housing, is used on water-heaters. The time switch opens the main heater circuit during predetermined peak-load conditions

A combination watt-hour meter and demand meter is used to indicate the maximum demand in addition to recording watt-hours. This device is made in two different constructions The thermal type combines in a single housing a thermal-demand indicator with a single or multistator watt-hour meter. The mechanical, or integrated-demand, type consists of a single or multistator watt-hour meter equipped with a demand register instead of a conventional watt-hour register The demand register has, in addition to the watt-hour register parts, a timing means and a mechanism for integrating the energy consumed over the demand interval.

A watt-hour meter with a contact device is used for measurement of demand, particularly when large blocks of energy are involved A demand meter, located externally to the watt-hour meter, is actuated through a contact device contained in the watt-hour meter and geared to the rotor shaft.

[C.R S]

Bibliography: American Standards Association, *Code for Electricity Meters*, C12, 1941; supplement C12a, 1947; Edison Electric Institute, *AEIC-EEI-NEMA Standards for Watthour Meters*, Publ MSJ-10; Edison Electric Institute, *Electrical Metermen's Handbook*, 6th ed., 1950; F. K. Harris *Electrical Measurements*, 1952; I. F. Kinnard, *Applied Electrical Measurements*, 1956, A. E. Knowlton, *Electric Power Metering*, 1934

Wattmeter

An instrument that measures electric power. For a complete discussion of the power in various types of electric circuits, *see* ELECTRIC POWER MEASUREMENT.

A variety of wattmeters is available to measure the power in ac circuits. They are generally classified by names descriptive of their operating principles. Determination of power in dc circuits is almost always done by separate measurements of voltage and current. However, some of the instruments described will also function in dc circuits, if desired

Electrodynamic wattmeter. Probably the most useful instrument in the measurement of ac power at commercial frequencies is the indicating (deflecting) electrodynamic wattmeter. It is similar in principle to the double-coil dc ammeter or voltmeter in that it depends on the interaction of the fields of two sets of coils, one fixed and the other movable The moving coil is suspended, or pivoted, so that it is free to rotate through a limited angle about an axis perpendicular to that of the fixed coils. As a single-phase wattmeter, the moving (potential) coil, usually constructed of fine wire, carries a current proportional to the voltage applied to the measured circuit, and the fixed (current) coil carries the load current. This arrangement of coils is due to the practical necessity of designing current coils of relatively heavy conductors to carry large values of current. The potential coils can be lighter because the operating current is limited to low values

If i_1 is the instantaneous current in the potential coil and i_2 is the instantaneous current in the current coil, and there is no iron or disturbing magnetic field due to current in neighboring conductors, then

$$\text{Instantaneous torque} = k\,(i_1 i_2)$$

Since i_1 is proportional to the instantaneous voltage e across the circuit,

$$\text{Instantaneous torque} = k\,(e i_2)$$

The moving system however, is designed with sufficient inertia that it is unable to follow the rapid alternations of the alternating current, but it will rotate, opposed by light zeroing springs, to a position corresponding to the average torque, which is proportional to the average power being supplied A modern wattmeter of the electrodynamic type is shown in Fig. 1

To avoid a loss of power in the instrument due to the current flowing in the potential coil, a fine wire coil is wound with the current coil and connected in series with the potential coil. Its effect is to cancel the magnetic effect of the potential coil current from the field of the current coils. If this compensation is correct, with the load circuit open the instrument will read zero.

The presence of inductance in the potential circuit would normally introduce phase displacement

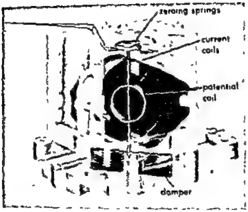


Fig 1 Single-element electrodynamometer wattmeter (Weston Instruments, Division of Daystrom, Inc.)

between voltage and current, which is theoretically inadmissible Simple tuning with capacity would introduce frequency error, but capacity in series with the moving coil and shunted by noninductive resistance is found to reduce the net reactance to a tolerably low value.

Other disturbing effects are those of ambient temperature, eddy currents in metal structure close to the moving coils, transformer effect due to mutual induction betw[illegible]

[illegible] ts are taken

[illegible]racy within its range and is commonly used as a laboratory standard It has the great advantage of being usable as a transfer instrument, that is, it can be calibrated with direct current and used for ac measurements. It is chiefly useful at power frequencies In modern practice, the ranges of voltage are limited to about 150 volts, though external multipliers to extend the voltage range may be used Maximum current is usually about 20 amps Shunting of current coils is not recommended The preferred method of extending the range of both current and voltage is through current and potential transformers by which the voltages and current in the load circuit are reduced to nominal values by definite transformation ratios without introducing appreciable phase errors, or, if such errors are introduced, their values are known and may be accounted for See INSTRUMENT TRANSFORMER

Thermal wattmeter. The thermal wattmeter is a versatile means of power measurement, since its operation is based on the heating effect of current, the I^2R relationship. It is applicable to both direct current and alternating current and is usable at frequencies up into the radio range without regard to waveform Various electronic wattmeters for audio and radio power measurements are largely based on this principle Thermocouples are commonly used as transducers, and the derived voltage is available for the common types of dc indicators and for potentiometers of either the manual or automatic self balancing type. A limitation is that the thermal wattmeter must be calibrated against another wattmeter. Also, it is relatively slow in response and may be affected by temperature.

A typical system used in commercial power measurements is shown in Fig. 2. The transducer elements are placed in the semicylindrical structure shown, which is divided into two compartments designed to limit cooling by convection. Each compartment contains a resistance-thermocouple assembly, which is made up of a series of thermocouples having the cold junctions *a* mounted on short metal posts and the hot junctions *b* suspended in air. The mounting posts are electrically insulated from, but in close thermal association with the base of the [illegible]

effect but also an ohmic resistance, which may be heated by the load current. Thus, in each of these identical elements, a thermoelectric potential is developed with a polarity dependent on the arrangement of the thermoelectric pairs and a magnitude proportional to the square of the current flowing The load current is applied to the primary of a current transformer with two identical secondary windings I_1 and I_2, which are differentially connected to the transducer elements. The line voltage is connected in series with a resistance to a potential transformer E whose secondary voltage is tapped into the midpoints of secondaries of the current transformer. If the circuit elements are symmetrical, the temperature rise of the free ends of

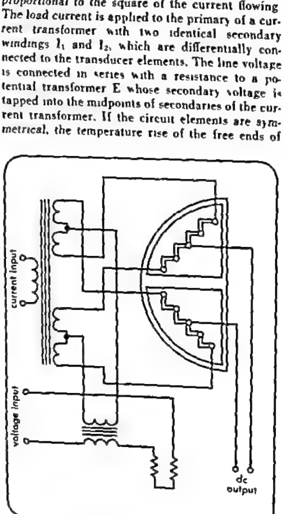


Fig 2 Thermal wattmeter circuit

the thermocouples is proportional to the power dissipated in each element or to the squares of the corresponding currents These values are different, since they depend on sum and difference terms, and thus the value and polarities of the thermoelectric elements are such that the dc potential existing between the midpoints of the two elements is proportional to the respective temperatures and, therefore, to the squares of the respective currents and, in turn, to the power in the ac circuit.

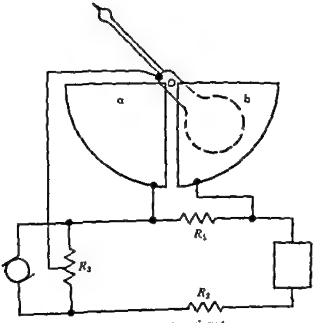
As with other methods, the range of voltage and current is extended through the use of additional current and potential transformers in the circuit external to the equipment shown in the figure.

Electrostatic wattmeter. The elementary quadrant electrometer has been adapted for power measurements A schematic diagram of this instrument is shown in Fig 3. The mechanism consists of two quadrants a and b charged by the voltage drop across a noninductive shunt resistance R_1 through which the load current passes. The line voltage is applied between the moving vane and one of the quadrants. In the circuit shown, the voltage to the moving vane is taken from the voltage divider R_3. This, in combination with the series resistance R_2, is a common means for providing compensation for power losses in the shunt. All resistances are noninductive The deflection θ of the indicator is proportional to the average power, that is,

$$K\theta = \frac{1}{T}\int_0^T ei\,dt$$

where T is the period of the alternating wave.

The method is unique among others in use in that it is a voltage method rather than a more usual current method. It has the advantages of (1) the possibility of wide ranges of measured current and voltage, (2) readings essentially unaffected by ambient conditions, and (3) measurements generally free of errors due to frequency or waveform



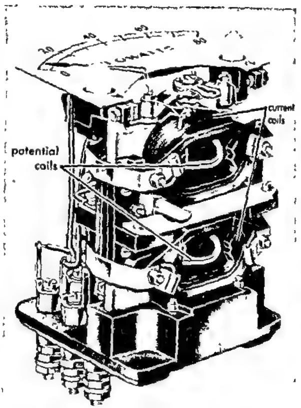


Fig 4 Two-element, polyphase electrodynamometer wattmeter (*Weston Instruments, Division of Daystrom, Inc*)

Its disadvantages are that it is relatively insensitive being limited by the voltage drop permitted across the shunt and that it requires careful screening to eliminate all possible effects of charged bodies in the vicinity. It also must be calibrated at the voltages at which it is used Moreover, the movement is generally low in torque and weight and hence relatively delicate mechanically.

This instrument has found its chief use in the laboratory for standardizing purposes and in capacity testing where small values of power, low power factors, and high voltages are involved

Polyphase wattmeter. The instruments thus far considered are designed for single-phase power measurement. In polyphase circuits, the total power is the algebraic sum of the power in each phase This summation is assisted by simple modifications of single-phase instruments.

For example an electrodynamic wattmeter may contain a second coil system similar to the coils of a single-phase meter, with the second potential coil on the same shaft as the potential coil in the first system The two systems are mounted in the same case and are designed to have matched characteristics but care is taken that there is no magnetic interaction between them (see Fig. 4). The deflection is proportional to the sum of the torques of the two elements; thus, total power is read from the instrument scale Electrical connections are the same as for two single-phase wattmeters

The thermal wattmeter is also commonly made in a two element system, the second element being identical with the first and ... nsducer networks being connected ... ed d

put is the algebraic sum of the ac power in the two systems. [C.A.M.]

Bibliography: C. V. Drysdale and A. C. Jolley, *Electrical Measuring Instruments*, pt. 1, 2d ed., 1952, F. K. Harris, *Electrical Measurements*, 1952, F. A. Laws, *Electrical Measurements*, 2d ed., 1938, F. E. Terman and J. M. Pettit, *Electronic Measurements*, 2d ed., 1952.

Wave (capillary)

Capillary waves, or ripples, are waves at the interface between two fluids, in which the principal restoring force is controlled by surface tension. Ripples generated by wind at the interface between air and water on oceans and lakes are of importance to the friction of air flowing over water, and to the reflection and scattering of electromagnetic and sound waves. See SURFACE TENSION.

The formulas relating the phase velocity c and frequency f to the wavelength λ of low amplitude, sinusoidal waves are, in the absence of wind forces,

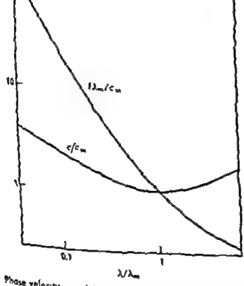
$$c = c_m\left(\frac{\lambda}{2\lambda_m} + \frac{\lambda_m}{2\lambda}\right)^{1/2} \qquad f = \frac{c}{\lambda}$$

where

$$c_m = \left(2\frac{\rho_2 - \rho_1}{\rho_2 + \rho_1}\right)^{1/2}\left(\frac{gT}{\rho_2 - \rho_1}\right)^{1/4}$$

$$\lambda_m = 2\pi\left[\frac{T}{(\rho_2 - \rho_1)g}\right]^{1/2}$$

T is the surface tension, ρ_1 and ρ_2 are the densities of upper and lower fluids, respectively, and g is the acceleration of gravity. For air over water at 15°C, c_m is 23 cm/sec, λ_m is 1.7 cm. The phase velocity is a minimum at $c = c_m$ when $\lambda = \lambda_m$ (see figure). Shorter waves are ripples, longer waves are gravity waves. See OCEAN WAVES.



Phase velocity c and frequency f of ripples and gravity waves as functions of wavelength plotted on logarithmic scales. λ_m is wavelength of minimum phase velocity c_m.

In nature, ripples are observed to grow rapidly when the wind blows and to die away rapidly when the wind stops. When the water surface is uncontaminated, ripples die away to e^{-1} of their original amplitude in a time $t_0 = \lambda^2/(8\pi^2\nu)$, where ν is the kinematic viscosity of water. For $\lambda = 1.7$ cm and $\nu = .01$ cm²/sec, $t_0 = 3.8$ sec. When the water surface is contaminated, as by an oil film or other surface active agent, ripples are damped still more rapidly because the contaminated surface acts as an inextensible film against which the water motions due to the ripples must rub. For a perfectly inextensible film, t_0 is equal to $(\pi^2\nu f)^{-1/2}$. For the example treated above, t_0 becomes 0.86 sec. For moderate winds, the increased damping of ripples almost completely inhibits their growth; the surface appears smooth and is called a slick. It has been observed that even gravity waves grow at an inappreciable rate under such conditions, the interpretation here is that the fine scale of roughness presented to the wind by a rippled surface is necessary for the formation of gravity waves. On the other hand, ripples have been observed to be formed in the absence of wind, by momentary nonlinear interactions of steep gravity waves, consequently the formation of both gravity waves and ripples is an interconnected process.

The surface profile of ripples of large amplitude which move without change of form has been calculated by G. D. Crapper. The profile changes from sinusoidal to one with sharper troughs than crests as the amplitude increases. [C.S.C.]

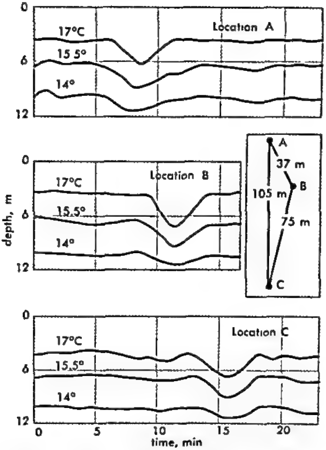
Bibliography: G. D. Crapper, An exact solution for progressive capillary waves of arbitrary amplitude, *J. Fluid Mech.*, 2:532–540, 1957; A. H. Schooley, Profiles of wind-water waves in the capillary gravity transition region, *J. Marine Research*, 16(2):100–108, 1958.

Wave (internal)

Internal waves are wave motions of stably stratified fluids in which the maximum vertical motion takes place below the surface of the fluid. The restoring force is mainly due to gravity; when light fluid from upper layers is depressed into the heavy lower layers, buoyancy forces tend to return the layers to their equilibrium positions. Internal waves have been found in the atmosphere as lee waves (waves in the wind stream downwind from a mountain) and as waves propagated along an inversion layer (a layer of very stable air). In the oceans, internal oscillations have been observed wherever suitable measurements have been made, but it is not completely certain that all of these oscillations are ...

... minutes and the longest periods being many days. At a number of locations in the oceans, internal tides, or internal waves having the same periodicity as oceanic tides, are prominent.

The vertical distribution of motions of internal waves depends on the vertical gradient of density



Diagrams showing passage of internal wave in the ocean off Mission Beach, California Curves on each diagram indicate depth of certain isotherms recorded at locations A, B, and C as functions of time. Prominent trough recorded at successively later times at the three locations represents a solitary internal wave trough traveling at a speed of 23 cm/sec Average depth to sea floor was 20 m Insert shows relative location of recorders in horizontal plan

in the fluid and the frequency of the generating forces There is a simple density distribution which is illustrative: the fluid consists of two homogeneous layers, a lighter one on top of a heavier one, such as kerosine over water. The internal waves in this system are sometimes called boundary waves because the maximum vertical motion occurs at the discontinuity of density at the boundary Let the thickness of the layers be h_1 and h_2, let g be the acceleration of gravity and let $\delta\rho/\rho$ be the fractional change of density across the boundary (In the ocean $\delta\rho/\rho$ is of order 0.1% and one may neglect squares of this small quantity) Then the phase velocity of internal waves of wavelength long compared to $h_1 + h_2$ is

$$[(gh_1h_2)(h_1 + h_2)(\delta\rho/\rho)]^{1/2}$$

This is to be compared with

$$[g(h_1 + h_2)]^{1/2}$$

for surface waves of great length. Because of the factor $(\delta\rho/\rho)^{1/2}$, the internal waves move at a slow speed, of the order of a few knots in the deep oceans. The effect of the rotation of the earth is to increase the phase velocity of waves having periods long enough to approach one pendulum day

When there is a continuous distribution of density in the fluid, internal waves are only possible for frequencies lower than

$$(2\pi)^{-1}\left[g\frac{d}{dz}(\ln\rho)\right]^{1/2}$$

where $$\frac{d}{dz}(\ln\rho)$$

is the maximum downward rate of increase of the logarithm of density. At any frequency lower than this limit there is an infinity of possible modes of internal waves. In the first mode, the vertical motion has a single maximum somewhere in the body of the fluid. In the second mode there are two such maxima (180° out of phase) with a node between and so on.

Internal waves are thought to be generated in the sea by air pressure variations and by interaction with low-frequency surface waves in water of variable depth. Occasionally boundary waves have been excited by slow ships where a thin layer of fresh water overlies salt water. This phenomenon causes increased drag for the ships. It has been detected in the Northern Atlantic where melting ice creates such conditions. [csc]

Bibliography: J. E. Fjeldstad, *Interne Wellen*, Geofys. Publ. 10, no. 6, 1933; P. Groen, Contribution to the theory of internal waves, *Mededeel Verh Kgl. Ned. Meteorol. Inst.*, ser. B, vol 2, no 1, 1948, R. S. Scorer, *Natural Aerodynamics*, 1958, H. U. Sverdrup, M. W. Johnson, and R. H Fleming, *The Oceans*, 1942.

Wave (physics)

The general term applied to the description of a disturbance which propagates from one point in a medium to other points without giving the medium as a whole any permanent displacement.

Waves are generally described in terms of their amplitude, and how the amplitude varies with both space and time. The actual description of the wave amplitude involves a solution of the wave equation and the particular boundary conditions for the case being studied In the cases most often considered the wave equation is simplified to a second order, linear, partial differential equation This equation for a one-dimensional space coordinate x is

$$\frac{\partial^2\phi}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}$$

where ϕ is the amplitude, c is the wave velocity, and t is the time coordinate. The generalized solutions of this equation are of the form

$$\phi = F(x - ct) + G(x + ct)$$

The first term indicates a wave traveling in the positive x direction at a velocity c and the second term a wave traveling in the negative x direction The functions F and G are determined by the particular properties of the boundary conditions of the problem In the one-dim al case these are

usually sine or cosine waves The velocity of propagation c is proportional to the square root of the ratio of the elastic to the inertial constants of the medium. *See* SINE WAVE; WAVE EQUATION; WAVE MOTION

Acoustic waves, or sound waves, are a particular kind of the general class of elastic waves Elastic waves are propagated in media having two properties, inertia and elasticity. Elasticity of the medium is required in order to provide a force which tends to restore a displaced particle of the medium to its original position. Inertia is required to enable the displaced particle to transfer momentum to an adjoining particle A shear wave, or rotational wave, is a wave in an elastic medium which causes an element of the medium to change its shape without a change of volume.

Electromagnetic waves, for example, light waves and radio waves, are not elastic waves and therefore can travel through a vacuum. The velocity of the wave depends on the medium through which the wave travels, but in a vacuum it is a constant, c, approximately equal to 3×10^8 m/sec *See* ELECTROMAGNETIC WAVE; *see also* SHOCK WAVE; STANDING WAVE; STATIONARY WAVE. [W J C]

Wave equation

This term refers to the differential equation of wave motion, which can be applied to electromagnetic waves, sound waves, and other elastic waves. It can also refer to Schrödinger's wave equation of quantum theory This article is concerned with the wave equation as used in electromagnetic theory For applications of the wave equation in acoustics, *see* WAVE MOTION For information on Schrodinger's equation, *see* QUANTUM THEORY, NONRELATIVISTIC.

The differential equation for the propagation of electromagnetic fields may be written, in vector notation,

$$\nabla^2\mathbf{B} = \mu\gamma\frac{\partial\mathbf{B}}{\partial t} + \mu\epsilon\frac{\partial^2\mathbf{B}}{\partial t^2} \qquad (1)$$

where γ is the conductivity, μ the permeability, ϵ the capacitivity, or dielectric constant, and B the magnetic flux density An identical equation holds for the electric intensity E. In rectangular components, the wave equation separates into three equations of the form

$$\frac{\partial^2 B_u}{\partial x^2} + \frac{\partial^2 B_u}{\partial y^2} + \frac{\partial^2 B_u}{\partial z^2} = \mu\gamma\frac{\partial B_u}{\partial t} + \mu\epsilon\frac{\partial^2 B_u}{\partial t^2} \qquad (2)$$

where B_u stands for B_x, B_y, or B_z For simple solutions of Eq (2), *see* ELECTROMAGNETIC RADIATION, *see also* MAXWELL'S EQUATIONS When the fields vary sinusoidally with time, it is convenient to write Eq (1) in terms of phasors, which are complex numbers such that, when multiplied by $e^{j\omega t}$, the real part of the product gives the amplitude, phase, and time dependence. The first and second time derivatives in Eq (1) are then replaced by $j\omega$ and $-\omega^2$, respectively, and it takes the form

$$\nabla^2\check{\mathbf{B}} = (j\omega\mu\gamma - \omega^2\mu\epsilon)\check{\mathbf{B}} = \check{\Gamma}^2\check{\mathbf{B}} \qquad (3)$$

The propagation constant $\check{\Gamma}$ is complex in a conductor, but is a pure imaginary number in a perfect insulator. Its real part α is called the attenuation constant, and its imaginary part β or k, the phase constant or wave number. In terms of wavelength λ in meters, frequency ω in radians per second, frequency ν in cycles per second, and velocity v in meters per second, it is written

$$\beta = \frac{2\pi}{\lambda} = \frac{\omega}{v} = \frac{2\pi\nu}{v} \qquad (4)$$

Plane-wave solutions. When Laplace's operator ∇^2 is applied to a vector, only the rectangular unit vectors **i**, **j**, and **k** have everywhere the same magnitude and direction and may be taken out from under the operator Therefore, the simplest solution of Eqs (1) and (2) is that for a plane-polarized unbounded wave in a nonconducting isotropic medium The next simplest are for plane bounded waves The conductivity γ is taken to be zero for simplicity, but the general sinusoidal time dependent case may be treated by the use of phasor notation and the substitution of $\check{\Gamma}^2$ for $\omega^2\mu\epsilon$

It is readily verified by substitution in Eq. (2) that in a perfect insulator, solutions for a z-directed wave are

$$E_x = \frac{\partial U(x,y)}{\partial x} f[z - (\mu\epsilon)^{1/2}t] \qquad E_y = \frac{\partial U(x,y)}{\partial y} f[z - (\mu\epsilon)^{1/2}t] \qquad (5)$$

provided $U(x,y)$ satisfies Laplace's equation in two dimensions, which is

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0 \qquad (6)$$

By using Maxwell's equations, it can be shown that

$$(\mu\epsilon)^{-1/2}B_y = E_x = \frac{\partial U}{\partial x}f(z - vt) = \frac{\partial V}{\partial y}f(z - vt) \qquad (7)$$

$$(\mu\epsilon)^{-1/2}B_x = -E_y = -\frac{\partial U}{\partial y}f(z - vt) = \frac{\partial V}{\partial x}f(z - vt) \qquad (8)$$

Here $U(x,y)$ and $V(x,y)$ are conjugate functions such that

$$W = U + jV = f(x + jy) \qquad (9)$$

Thus, the powerful method of conformal transformations can be used to find plane-wave solutions in which the potential U, at some value of z, is specified on the surface of two or more cylindrical conductors of any shape This solves all uniform transmission line problems.

Another important plane-wave solution is that in which either E or B, but not both, has a z component These solutions describe wave propagation in straight conducting pipes of uniform cross section called wave guides The so-called transverse magnetic (TM) wave is obtained by finding a solution of Eq (2) for E_z which vanishes at the walls The

other components of the wave fields, which may not be in rectangular coordinates, are obtained by applying the last two of the four Maxwell equations. The transverse electric (TE) solution is found in the same way from a solution for B_z which is tangential to the walls. In these cases, it is found that progressive wave solutions occur only above a certain frequency, called cutoff, which depends on the dimensions of the pipe. The velocity of propagation of energy is less than that of an unbounded wave in the medium filling the tube, and the phase velocity is greater. *See* WAVE GUIDE.

Electric dipole radiation. The spherical wave is the next in importance to the plane wave. The most important and familiar of the spherical waves is that of the electric dipole of moment $M \cos \omega t$. When V is a function only of r, Laplace's operator becomes

$$\nabla^2 V = \frac{1}{r}\frac{\partial}{\partial r}\left(r^2 \frac{\partial V}{\partial r}\right)$$

The simplest sinusoidal solution of Eq. (1) in spherical polar coordinates when γ is zero and B is replaced by V is

$$V = Cr^{-1} \sin (\omega t - \beta r) \qquad (10)$$

where β^2 is $\omega^2\mu\epsilon$, and r is the radius vector from the origin so that

$$r^2 = \rho^2 + z^2 \quad \text{and} \quad \rho = r \sin \theta \qquad (11)$$

The field of a z-directed dipole is symmetrical about the z axis, and the following relation holds:

$$B_\phi = -\frac{\partial V}{\partial \rho} = C\left[\frac{\rho}{r^3} \sin (\omega t - \beta r) + \frac{\beta\rho}{r^2} \cos (\omega t - \beta r)\right]$$

where B_ϕ is the ϕ component of B and ϕ is the longitude angle around the z axis. The quantity $r \sin \theta$ replaces ρ and $\omega\mu M/4\pi$ replaces C, so that

$$B_\phi = -\frac{\omega\mu M \sin \theta}{4\pi r^2} [\beta r \cos (\omega t - \beta r) + \sin (\omega t - \beta r)] \qquad (12)$$

The electric fields of the wave are found by integration of the r and θ components of the fourth Maxwell equation.

$$E_r = -\frac{M \cos \theta}{2\pi\epsilon r^3} [\beta r \sin (\omega t - \beta r) - \cos (\omega t - \beta r)] \qquad (13)$$

$$E_\theta = \frac{M \sin \theta}{4\pi\epsilon r^3} [(1 - \beta^2 r^2) \cos (\omega t - \beta r) - \beta r \sin (\omega t - \beta r)] \qquad (14)$$

The verification that Eqs. (12) to (14) are solutions of the wave equation for an electric dipole is that, when r becomes very small, all βr and $\beta^2 r^2$ terms in Eqs. (13) and (14) disappear, leaving the well-known field expression for a static dipole multiplied by $\cos \omega t$. When r is very large, Eq (13) vanishes, while Eqs. (12) and (14) simplify to

$$B_\phi = (\mu\epsilon)^{1/2} E_\theta = \frac{\beta\omega\mu M \sin \theta}{4\pi r} \cos (\omega t - \beta r) \qquad (15)$$

This is often called the radiation field, and is strictly a radial wave with Poynting's vector pointed along r since E_r is zero (*see* POYNTING'S VECTOR). The intermediate βr terms in Eqs. (12) to (14) are not strictly radial waves and are often called the induction fields.

Standing waves. Solutions of the wave equation which may be written as a function of the coordinates multiplied by a trigonometric function of time are called standing waves. They occur, for example, when two waves which are polarized in the same plane and are of equal amplitude travel in opposite directions through the same medium The resultant fields are

$$E_x = 2E_0 \cos (\omega v^{-1} z) \cos \omega t$$
$$vB_y = 2E_0 \sin (\omega v^{-1} z) \sin \omega t \qquad (16)$$

Whenever ωz equals $\frac{1}{2}(2n + 1)\pi v$, where n is an integer, the electric field is zero at all times; whenever ωz equals $n\pi v$, the magnetic field vanishes These planes are called electric and magnetic nodes, respectively. Both types of nodes are spaced a half-wavelength apart The total energy is constant, but shifts back and forth between the electric and magnetic fields, being entirely electric when ωt is $n\pi$ and entirely magnetic when ωt is $(n + \frac{1}{2})\pi$.

Resonant lines and cavities. It is clear that if perfectly conducting planes are placed normal to z at any two of the electric nodes in Eq. (16), then the standing waves can continue oscillation indefinitely, except for losses in the medium and walls. The space between these planes is said to be resonant for these waves. The same is true for transmission lines When perfectly conducting planes are placed at the nodes in a wave guide normal to the walls, the intervening space is completely enclosed and constitutes a resonant cavity. These are used for frequency standards and to furnish the large electric fields in cyclotrons and other particle accelerators where charged particles are passed repeatedly through the cavity to attain very high velocities (*See* CAVITY RESONATOR.) [W.R.SM.]

Bibliography: P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, 2 vols, 1953, L Page and N I Adams, *Electrodynamics*, 1940; W. K. H. Panofsky and M Phillips, *Classical Electricity and Magnetism*, 1955; W R. Smythe, *Static and Dynamic Electricity*, 2d ed., 1950; J. A. Stratton, *Electromagnetic Theory*, 1941.

Wave guide

A device which constrains or guides the propagation of electromagnetic waves along a path defined by the physical construction of the guide. In a broad sense, devices like a pair of parallel wires and a coaxial cable can certainly be called wave guides When used in a more restricted sense, however, a wave guide usually means a metallic tube which can confine and guide the propagation of electromagnetic waves in the hollow space along the lengthwise direction of [illegible] tube. For reasons which will become clear i[illegible] discussion.

hollow wave guides of convenient sizes are best adapted to the transmission of microwaves.

The concept that hollow wave guides can transmit electromagnetic waves may seem strange to people who lean heavily on experiences with low-frequency waves. It will not appear so strange, however, if one thinks in terms of an analogy with sound waves going through pipes (for example, *pipe organs*). In view of the fact that a sound wave can transmit through a pipe only when its wavelength is comparable to or smaller than the size of the pipe, one would expect that a similar requirement should hold true for electromagnetic waves. Indeed, if the frequency of an electromagnetic wave is high enough that the wavelength is comparable to or smaller than the wave-guide dimension, then wave transmission through the hollow wave guide becomes possible.

Although hollow wave guides and coaxial cables are the commonest in application, some other types of wave guides are also occasionally used. A single conductor (called a G-string) is sometimes used as a wave guide. Another wave guide takes the form of a flat conducting strip having a certain spacing from a ground plane, known as a microwave strip. Still another example is found in a dielectric rod.

Maxwell's equations. The most basic approach to the understanding and analysis of the behavior of electromagnetic waves in any wave guide is obtained by the application of Maxwell's equations to a given physical situation. These are a set of partial differential equations relating the quantities electric intensity **E**, electric induction **D**, magnetic intensity **H**, and magnetic induction **B**. Each quantity is regarded as a vector, having a direction as well as a magnitude which is a function of space coordinates and time. In the ordinary case of a homogeneous isotropic medium, electric induction is proportional to electric intensity. The constant of proportionality is known as the dielectric constant ϵ. Similarly, magnetic induction is related to magnetic intensity by a constant known as the permeability μ. See MAXWELL'S EQUATIONS.

A general solution of Maxwell's equations leads to a wave equation which points definitely to the possible existence of electromagnetic waves in the medium. Thus, for a dielectric medium of infinite extent, all solutions of the wave equation are equally admissible. One common characteristic for all these waves is that there is a velocity of propagation which is completely determined by the dielectric and permeability constants of the medium. See WAVE EQUATION.

Any particular solution for a realizable guided wave must obey certain boundary conditions imposed by the physical situation. The walls of a hollow wave guide are almost always made of a highly conducting metal like copper, brass, or aluminum. The electrical conductivity of such materials, while always finite, is so high that it can be considered to be infinite for the present considerations. It is well known that no electric intensity can exist inside a perfect conductor. On the basis of this fact and the application of Maxwell's equations, the boundary conditions for the electric and magnetic intensities at the interface between a perfect conductor and a dielectric (usually air) in a hollow wave guide turn out to be that both the tangential component of **E** and the normal component of **H** are zero at the boundary.

Transmission modes. Consider a hollow wave guide with a given cross section which is uniform throughout its entire length. As a result of the application of these boundary conditions to the wave equation, it can be shown that only certain unique patterns for the distribution of **E** and **H** (taken together) can exist in the wave guide. Each unique pattern of the field distribution is called a mode. There are two types of mode possible in a hollow wave guide. One type is called the transverse electric (TE) mode, in which **E** has only a component transverse (that is, perpendicular) to the direction of propagation, whereas the magnetic intensity **H** has both transversal and longitudinal components. The other type is called the transverse magnetic (TM) mode, in which the magnetic intensity has only a transverse component and the electric intensity has both components. Each type (TE or TM) of mode has an infinite number of submodes which have the common characteristics of the type to which they belong, but differ among themselves in the details of field distribution. Since it is known that the transverse electric and magnetic (TEM) mode is not possible in a hollow wave guide, any arbitrary electromagnetic wave inside such a wave guide can be considered as a linear superposition of all possible modes of both the TE and TM types.

Rectangular wave guides. The type of wave guide with a rectangular cross section is not only the commonest in use but also the simplest in theoretical analysis. It will be used here as a concrete example to illustrate various common properties of a wave guide. Consider a rectangular wave guide as shown in Fig. 1. The wave propagates

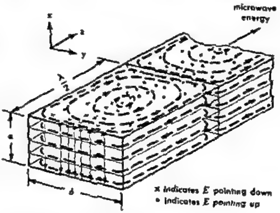


Fig. 1. Instantaneous field pattern for the TE_{01} wave in a rectangular wave guide. Solid lines indicate the electric intensity **E** and dashed lines the magnetic intensity **H**. (From MIT Radar School Staff, Principles of Radar, 1952)

along the z axis. The simplest and also the most commonly used mode is called the TE_{01} wave; its electric and magnetic intensities can be described by the following expressions satisfying the boundary conditions:

$$
\begin{aligned}
E_x &= A \sin\left(\frac{\pi}{b}y\right) \sin\omega\left(t - \frac{z}{v}\right) \\
H_y &= \frac{1}{\mu v} A \sin\left(\frac{\pi}{b}y\right) \sin\omega\left(t - \frac{z}{v}\right) \qquad (1) \\
H_z &= -\frac{\pi}{\mu\omega b} A \cos\left(\frac{\pi}{b}y\right) \cos\omega\left(t - \frac{z}{v}\right) \\
H_x &= E_y = E_z = 0
\end{aligned}
$$

where A = arbitrary constant depending upon the strength of wave excitation, $\omega = 2\pi \times$ frequency, t = time, v = velocity of propagation of the wave, and μ = permeability of the dielectric filling the wave guide. It is particularly important to note that the velocity of propagation v in a wave guide is different from that in an infinite space filled with the same dielectric material v_0. These two quantities are related for the TE_{01} wave by the equation:

$$v = \frac{v_0}{\sqrt{1 - (v_0/2fb)^2}} \qquad (2)$$

where f = frequency. By applying the simple formula $v = f\lambda$, it is seen that the wavelength in the wave guide is similarly related to the corresponding quantity in an infinite dielectric (λ_0) as (for the TE_{01} wave)

$$\lambda = \frac{\lambda_0}{\sqrt{1 - (\lambda_0/2b)^2}} \qquad (3)$$

The expressions in Eq. (1) reveal the wave nature of the field quantities through the sinusoidal functions of $\omega(t - z/v)$, which are a description of wave motion with velocity v along the z axis. On any cross section (constant z), these sinusoidal functions depend simply on time. Hence, each of the field quantities oscillates at the common frequency f and with an amplitude which varies with the field point (y in this particular case).

Phase velocity. Equation (2) represents the phase velocity of the guided wave because the quantity v is contained in $\omega(t - z/v)$, which is called the phase of wave propagation. It is seen that if $\lambda_0 < 2b$, the phase velocity in a wave guide is larger than that in an open space filled with the same dielectric material. Correspondingly, the wavelength in a wave guide is longer than that in an infinite dielectric medium, as indicated by Eq. (3). All waves with $\lambda_0 < 2b$ belong to the transmission region of the wave guide because only the waves in this region are allowed to pass through. When $\lambda_0 = 2b$, both v and λ are infinite, and when $\lambda_0 > 2b$, they are imaginary. It can be shown that waves are not allowed to propagate in a wave guide when $\lambda_0 \geqq 2b$. The critical value of $\lambda_c = 2b$ is known as the cutoff, or critical, wavelength.

Generalization to other TE waves. The preceding discussion can be carried through in a parallel fashion for the general case of any TE wave designated as TE_{nm}, where n and m are integers. Each TE_{nm} wave will have its characteristic field distribution, velocity, and wavelength. The expression for the wavelength in the wave guide, for example, is

$$\lambda = \lambda_0 \Bigg/ \sqrt{1 - \left(\frac{n\lambda_0}{2a}\right)^2 - \left(\frac{m\lambda_0}{2b}\right)^2} \qquad (4)$$

The corresponding cutoff wavelength is

$$\lambda_c = 2 \Bigg/ \sqrt{\left(\frac{n}{a}\right)^2 + \left(\frac{m}{b}\right)^2} \qquad (5)$$

which reduces to $\lambda_c = 2b$ for the special case of TE_{01} where $n = 0$ and $m = 1$. A TE_{00} wave ($n = 0$ and $m = 0$) would have an infinite cutoff wavelength which is characteristic of a principal mode, if such were possible. Actually, the solution of the field equation shows that the principal mode cannot exist in a hollow metallic wave guide. Equation (5) further shows that any other values of n and m in TE_{nm} would lead to a cutoff wavelength shorter than that of TE_{01}, which is called the dominant mode for this reason.

TM waves. The TM waves can be treated in almost exactly the same manner as the TE waves. Although the field distributions in the two cases are completely different, the wavelengths (similarly, the velocities of propagation) in the guide for both cases obey Eqs. (4) and (5), except that neither n nor m can become zero for a TM_{nm} wave.

Designation of rectangular wave guides. It is highly desirable and customary for practical use to excite only the dominant mode (TE_{01}) in a rectangular wave guide. This means that a wave guide of a given size is useful for this one mode operation only for a certain range of free-space wavelengths (or frequencies). Four wave guides of certain dimensions and their corresponding wavelength bands have been conventionally designated by the letters S, C, X, and K, as shown in the table.

Band designation for rectangular wave guides

Band	Dimension, cm*	λ_0, cm
S	7.62 × 2.54	8.9–10.5
C	3.48 × 1.58	3.7– 5.1
X	2.54 × 1.27	3.0– 3.5
K	1.06 × 0.43	1.2– 1.5

Joints, bends, and junctions.

To achieve certain effects, wave guides are frequently joined together, bent or twisted, or formed into networks known as junctions.

Joints. Wave guides are often joined together under various conditions. Identical wave-guide terminals may be connected from end to end for extension or for inserting circuit elements. This can be done by providing a flange at each end of the wave guide and then butting the two flat surfaces of the flanges together so that the wave-guide ends form a contact joint. Sometimes, to

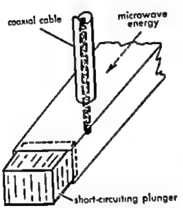


Fig 2 Coupling between a wave guide and a coaxial cable

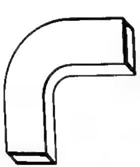


Fig 3 H plane 90° bend for a rectangular wave guide

alleviate the difficulty of making good contact, a choke joint is operated. This is based on the principle that an artificially created current nodal line at the joint will make the physical contact of the wave-guide ends unnecessary

Another type of wave-guide joint is required when a coaxial cable is to be joined to a hollow wave guide Here, a transformation of the wave is to take place between the coaxial mode and, for instance, the TE_{01} mode of a rectangular wave guide. In this case coaxial cable is usually led into the blocked end (or into a tunable short-circuiting plunger) of the regular wave guide and the center conductor is extended to touch the opposite face so that the extended wire will act as an antenna for the excitation of the wave in the rectangular guide (see Fig 2).

Bends and twists Wave-guide bends are often used to change the direction of the wave guide by a desired angle. There are two types of 90° bend—one in the H plane and the other in the E-plane A 90° bend in the H plane is shown in Fig 3 Twists are used to change the plane of polarization by a desired angle while maintaining the direction of the wave guide. A 90° twist is shown in Fig 4

Junctions The term junction is used to denote a network of wave guides which are joined in a specified manner to give certain desired properties to the whole network Some common examples are a T junction, a directional coupler, and a magic-tee junction.

... network with three wave-guide ... T. With ... ys of arranging a symmetrical T junction: either all three broad sides are in one plane or two broad sides are in one plane and the third in a perpendicular plane; the latter arrangement is shown in Fig. 5. In any case, when a microwave is incident to one wave-guide terminal, the incoming power will be divided equally between the remaining two wave guides, and the two branch waves will go along their separate directions.

A directional coupler is a network of four wave-guide terminals When suitably arranged, it has the property that for each terminal there is another with which it does not interact in any way. A special directional coupler known as the magic tee is described here For a discussion of the two-hole interference type of directional coupler, *see* DIRECTIONAL COUPLER.

A magic-tee junction is shown in Fig. 6 This network has a plane of symmetry as indicated. Consider only the dominant mode TE_{01} for all the wave guides. The mode in wave guide A is always symmetrical to the plane of symmetry, while the mode in wave guide C is always antisymmetrical. The symmetrical mode in A can excite symmetric

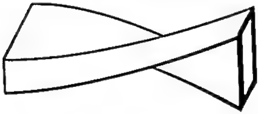


Fig 4 A 90° twist for a rectangular wave guide

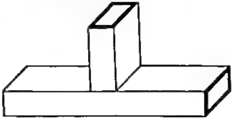


Fig 5 A T junction.

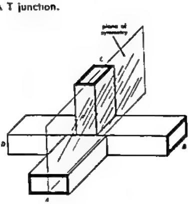


Fig 6 A magic-tee junction

modes in B and D but not in C. The antisymmetric mode in C can excite antisymmetric modes in B and D but not in A. A superposition of simultaneous excitations in A and C of equal strength must necessarily cause the intensity in either B or D to vanish because, if it is exactly in phase for one of them, it must be exactly out of phase for the other. Hence, in addition to the isolation between A and C, there is an isolation between B and D. These rather intricate properties are used in microwave circuits in a variety of ways, including the direct measurement of incident and reflected intensities.

Wave-guide shapes. Next to the rectangular type, metallic wave guides with circular cross sections are most frequently used. Because of its circular cylindrical symmetry, this type of wave guide can sustain waves of all kinds of field polarization (linear, circular, and elliptic) in many ways which are not possible with the rectangular type. However, the general notions of transmission modes, propagation velocities, cutoff wavelengths, etc., are just as applicable to circular wave guides as they are to rectangular wave guides, except that the mathematics used for circular shapes is somewhat more involved. Wave guides of elliptical cross section are almost never used on purpose, but they may represent results of deformation from the original circular wave guides. Detailed descriptions of wave behavior in elliptical wave guides are available in the literature.

Attenuation of hollow wave guides. In the preceding discussion, there has been an implicit assumption that the attenuation of a wave by the wave guide is negligible. While this assumption is for most purposes quite justifiable, it may be well to mention very briefly some general characteristics about the attenuation. As a wave passes through a wave guide (say in the z direction), its power becomes reduced exponentially according to the equation

$$P_z = P_0 e^{-\alpha z} \qquad (6)$$

where P_0 is the power at $z = 0$, P_z is the power at a distance z, and α is called the attenuation constant of the wave guide. The quantity α is a measure of the lossy character of the wave guide in absorbing the wave energy because of the currents set up on the inner walls of the wave guide and because of dielectric losses, if any. The better the wave guide is as a conductor, the smaller α becomes. The quantity α usually gets smaller with decreasing frequency (longer free-space wavelength), and it also changes with the mode of the excited wave.

Hollow wave guides are usually made of copper or brass, which give an attenuation constant somewhat less than 1 db/m in the X-band. It is seldom worthwhile to coat the inside surfaces with silver, which often reduces the attenuation by some amount. A hollow wave guide is usually much less lossy than a coaxial cable of comparable size, particularly when the latter is internally supported by lossy dielectrics.

Coaxial cable. A coaxial cable consists of a hollow cylindrical conductor coaxially placed with respect to an inner cylindrical conductor. It is very extensively used as a wave guide, though seldom so called, for very high as well as very low frequencies. The adaptability of a coaxial cable to a wide range of frequencies (including zero) is due to the existence of a principal mode, which is in fact a TEM wave. The analysis of the wave propagation in a coaxial cable is simple and is analogous to that of two parallel wires or plane plates. The velocity of propagation in such systems is the same as in infinite space filled with the same dielectric material. In common with a hollow wave guide, coaxial cable has the advantage that the outside conductor acts as a shield against external electrical interference. However, it does not have the ability possessed by hollow wave guides of filtering low frequencies, when such filtering is desired. Also, for very high frequencies, the attenuation of a coaxial cable is apt to be higher than that of a comparable hollow wave guide because of higher losses due to either the dielectric medium or the supports. *See* TRANSMISSION LINES

Other special wave guides. A single conductor, often thinly coated with a dielectric substance, can act as a wave guide under the nickname of G-string. A wave, usually of the TM mode, is guided along the surface of the conductor. The transmission characteristics are quite favorable in the frequency range of 80 to 300 Mc/sec.

A special wave guide adaptable to microwave circuit wiring has been developed under the name of a microwave strip. It consists of a flat conducting strip separated from the ground plane by a dielectric layer. Its chief advantage lies in the convenience of fabrication by printed circuit techniques.

Dielectric rods not involving any conductor can guide very high frequency (VHF) waves quite successfully. The propagating wave is partly inside the dielectric and partly outside. Such dielectric wave guides can be used for short-distance transmission. *See* MICROWAVE TRANSMISSION LINES; TRANSMISSION THEORY AND METHODS. [C.K.J.]

Bibliography: *See* MICROWAVE.

Wave mechanics

The modern theory of matter holding that elementary particles (such as electrons, protons, and neutrons) have wavelike properties. In 1924 Louis de Broglie postulated that the same wave-corpuscle duality which was then known to exist in the case of light might also occur in matter; this hypothesis was subsequently verified experimentally (see DE BROGLIE WAVELENGTH). With contributions by the mathematical physicists Erwin Schrödinger, Max Born, Werner Heisenberg, P. A. M. Dirac, and others, this theory of matter has become the highly successful quantum mechanics of the present day. For a discussion of wave mechanics and wave particle duality, *see* QUANTUM MECHANICS. [E.G.]

Wave motion

The process by which a disturbance at one point in space is propagated to another point more remote from the source with no net transport of the material of the medium itself. For example, sound is a form of wave motion, wind is not. Wave motion can occur only in a medium in which energy can be stored in both kinetic and potential form. In a mechanical medium, kinetic energy results from inertia and is stored in the motion of the molecules, while potential energy results from elasticity and is stored in the displacement of the molecules. In an electromagnetic medium, kinetic energy is stored in the magnetic field and potential energy in the electric field.

In a free traveling wave (as distinguished from a stationary or standing wave) one part of the medium disturbs an adjacent part, thereby imparting energy to it. This portion of the medium, in turn, disturbs another part, thereby causing a flow of energy in a given direction away from the source. More technically, wave propagation is the result of kinetic energy at one point being transferred into potential energy at an adjacent point, and vice versa. The rate of travel of the disturbance, or velocity of propagation, is determined by the constants of the medium. A stationary wave is the combination of two oppositely traveling waves of the same frequency and strength so that no net transfer of energy away from the source takes place. A standing wave is the same but with the returning wave (toward the source) being of lesser strength than the outwardly traveling wave so that a net transfer of energy away from the source does take place. *See* STANDING WAVE; STATIONARY WAVE.

Wave motion can occur in vacuo (electromagnetic waves), in gases (sound waves), in liquids (hydrodynamic waves), and in solids (vibration waves). Electro- ...

... the dielectric constant is not infinitely great. By current usage, elastic waves propagated in gases, liquids, and solids, regardless of whether hearable or not, are called acoustic waves.

In this article, it is shown that wave motion can exist in any continuous medium in which potential and kinetic energy can be stored. The wave equation is derived, and solutions for electromagnetic and mechanical media are given. It is shown that longitudinal waves may exist in all mechanical media, and that transverse waves may exist in electromagnetic and solid mechanical media. In all cases, viscosity or heat losses in mechanical media, and resistance and magnetic losses in electromagnetic media are neglected. For information which is related to, and supplements, the present discussion, *see* DIFFRACTION; ELECTROMAGNETIC RADIATION; HARMONIC MOTION; HUYGENS' PRINCIPLE; INTERFERENCE OF WAVES; LIGHT; MAXWELL'S EQUATIONS; REFRACTION OF WAVES; SUPERPOSITION, PRINCIPLE OF; VIBRATION; WAVE EQUATION; WAVE MOTION IN FLUIDS; WAVE MOTION IN LIQUIDS.

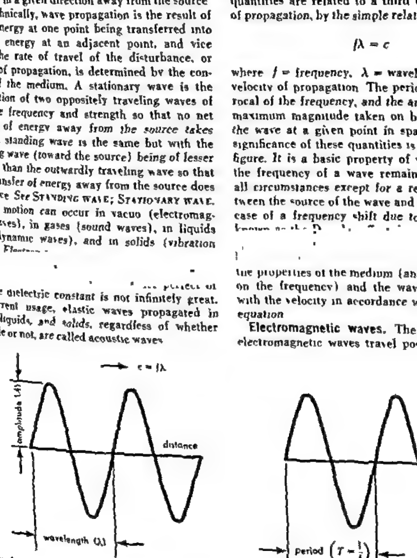
Fundamental relations. A wave is commonly referred to in terms of either its wavelength or its frequency. In any type of wave motion, these two quantities are related to a third quantity, velocity of propagation, by the simple relation

$$f\lambda = c$$

where f = frequency, λ = wavelength, and c = velocity of propagation. The period T is the reciprocal of the frequency, and the amplitude A is the maximum magnitude taken on by the variable of the wave at a given point in space. The physical significance of these quantities is illustrated in the figure. It is a basic property of wave motion that the frequency of a wave remains constant under all circumstances except for a relative motion between the source of the wave and the observer. The case of a frequency shift due to relative motion, known as the D ...

... the properties of the medium (and, sometimes, also on the frequency) and the wavelength will vary with the velocity in accordance with the preceding equation.

Electromagnetic waves. The media in which electromagnetic waves travel possess no elasticity



Relation between frequency, wavelength, and velocity in wave propagation

modes in B and D but not in C. The antisymmetric mode in C can excite antisymmetric modes in B and D but not in A. A superposition of simultaneous excitations in A and C of equal strength must necessarily cause the intensity in either B or D to vanish because, if it is exactly in phase for one of them, it must be exactly out of phase for the other. Hence, in addition to the isolation between A and C, there is an isolation between B and D. These rather intricate properties are used in microwave circuits in a variety of ways, including the direct measurement of incident and reflected intensities.

Wave-guide shapes. Next to the rectangular type, metallic wave guides with circular cross sections are most frequently used. Because of its circular cylindrical symmetry, this type of wave guide can sustain waves of all kinds of field polarization (linear, circular, and elliptic) in many ways which are not possible with the rectangular type. However, the general notions of transmission modes, propagation velocities, cutoff wavelengths, etc., are just as applicable to circular wave guides as they are to rectangular wave guides, except that the mathematics used for circular shapes is somewhat more involved. Wave guides of elliptical cross section are almost never used on purpose, but they may represent results of deformation from the original circular wave guides Detailed descriptions of wave behavior in elliptical wave guides are available in the literature.

Attenuation of hollow wave guides. In the preceding discussion, there has been an implicit assumption that the attenuation of a wave by the wave guide is negligible. While this assumption is for most purposes quite justifiable, it may be well to mention very briefly some general characteristics about the attenuation. As a wave passes through a wave guide (say in the z direction), its power becomes reduced exponentially according to the equation

$$P_z = P_0 e^{-\alpha z} \qquad (6)$$

where P_0 is the power at $z = 0$. P_z is the power at a distance z, and α is called the attenuation constant of the wave guide. The quantity α is a measure of the lossy character of the wave guide in absorbing the wave energy because of the currents set up on the inner walls of the wave guide and because of dielectric losses, if any. The better the wave guide is as a conductor, the smaller α becomes The quantity α usually gets smaller with decreasing frequency (longer free-space wavelength), and it also changes with the mode of the excited wave.

Hollow wave guides are usually made of copper or brass, which give an attenuation constant somewhat less than 1 db/m in the X-band. It is seldom worthwhile to coat the inside surfaces with silver, which often reduces the attenuation by some amount. A hollow wave guide is usually much less lossy than a coaxial cable of comparable size, particularly when the latter is internally supported by lossy dielectrics.

Coaxial cable. A coaxial cable consists of a hollow cylindrical conductor coaxially placed with respect to an inner cylindrical conductor. It is very extensively used as a wave guide, though seldom so called, for very high as well as very low frequencies. The adaptability of a coaxial cable to a wide range of frequencies (including zero) is due to the existence of a principal mode, which is in fact a TEM wave. The analysis of the wave propagation in a coaxial cable is simple and is analogous to that of two parallel wires or plane plates. The velocity of propagation in such systems is the same as in infinite space filled with the same dielectric material. In common with a hollow wave guide, coaxial cable has the advantage that the outside conductor acts as a shield against external electrical interference. However, it does not have the ability possessed by hollow wave guides of filtering low frequencies, when such filtering is desired Also, for very high frequencies, the attenuation of a coaxial cable is apt to be higher than that of a comparable hollow wave guide because of higher losses due to either the dielectric medium or the supports. *See* TRANSMISSION LINES

Other special wave guides. A single conductor, often thinly coated with a dielectric substance, can act as a wave guide under the nickname of G-string. A wave, usually of the TM mode, is guided along the surface of the conductor. The transmission characteristics are quite favorable in the frequency range of 80 to 300 Mc/sec.

A special wave guide adaptable to microwave circuit wiring has been developed under the name of a microwave strip. It consists of a flat conducting strip separated from the ground plane by a dielectric layer Its chief advantage lies in the convenience of fabrication by printed circuit techniques.

Dielectric rods not involving any conductor can guide very high frequency (VHF) waves quite successfully. The propagating wave is partly inside the dielectric and partly outside. Such dielectric wave guides can be used for short-distance transmission. *See* MICROWAVE TRANSMISSION LINES; TRANSMISSION THEORY AND METHODS. [C.K.J.]

Bibliography *See* MICROWAVE.

Wave mechanics

The modern theory of matter holding that elementary particles (such as electrons, protons, and neutrons) have wavelike properties. In 1924 Louis de Broglie postulated that the same wave-corpuscle duality which was then known to exist in the case of light might also occur in matter, this hypothesis was subsequently verified experimentally (*see* DE BROGLIE WAVELENGTH). With contributions by the mathematical physicists Erwin Schrödinger, Max Born, Werner Heisenberg, P. A. M. Dirac, and others, this theory of matter has become the highly successful quantum mechanics of the present day. For a discussion of wave mechanics and wave particle duality, *see* QUANTUM MECHANICS. [E.C.]

ormable membrane When this packet is squeezed, is found to have stiffness (or elasticity). Its mass quals the mass of the air molecules contained herein Also, the deformable membrane keeps the mass inside constant.

From this description, one sees that it is a relatively simple matter to describe the motion of such packet as a function of time The mass element s labeled in any convenient manner, the most common being by its location at any convenient time. This method is called the Lagrangian description of the motion

To derive the equations for wave motion in a gas, express the concepts of inertia by Newton's second law, of elasticity by the perfect gas law, and of the deformable packet by the conservation of mass law. Assume that the packet is rectangular in shape. To find the equation of motion, suppose that the box is situated in a medium where the pressure p changes in space at a space rate of

$$\operatorname{grad} p = \mathbf{i}\frac{\partial p}{\partial x} + \mathbf{j}\frac{\partial p}{\partial y} + \mathbf{k}\frac{\partial p}{\partial z} \tag{13}$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors in the x, y, and z directions respectively, and p is the pressure at a point.

The net force $\mathbf{f}$ acting to move the box in some direction is equal to the vector summation of the gradients in force across the three pairs of faces of the packet times the respective separations of these faces; in the positive direction,

$$\mathbf{f} = -\left[\mathbf{i}\left(\frac{\partial p}{\partial x}\Delta x\right)\Delta y\,\Delta z + \mathbf{j}\left(\frac{\partial p}{\partial y}\Delta y\right)\Delta x\,\Delta z + \mathbf{k}\left(\frac{\partial p}{\partial z}\Delta z\right)\Delta x\,\Delta y\right]$$
$$= -V' \operatorname{grad} p \tag{14}$$

where V' is the average volume of the packet The positive gradient causes an acceleration of the box in the negative direction of x

By Newton's second law, the force acting to accelerate the packet is

$$\mathbf{f} = \rho' V' \frac{\partial \mathbf{q}}{\partial t} \tag{15}$$

where $\mathbf{q}$ is the average vector velocity of the gas in the packet, ρ' the average density of the gas in the packet, and $\rho' V'$ the total (constant) mass of the gas in the packet In writing Eq (15), it has been assumed that the packet is never displaced from its equilibrium position a significant part of a wavelength of sound

Combining Eqs (14) and (15) yields

$$-\operatorname{grad} p = \rho_0 \frac{\partial \mathbf{q}}{\partial t} \tag{16}$$

In keeping with the approximation just stated it has been assumed that ρ' (the instantaneous density) does not appreciably deviate from the average density ρ_0. These approximations are acceptable provided the sound pressures are below about 100 dynes/cm² (10 newtons/m²).

Assuming adiabatic expansions and contractions of the gas, and that the gas is perfect,

$$\frac{dP}{P} = -\frac{\gamma\, dV}{V} \tag{17}$$

where P is the total pressure in the gas and γ is the *ratio of specific heat at constant pressure to specific heat at constant volume.*

Let

$$P = P_0 + p$$
$$V = V_0 + \tau \tag{18}$$

where p and τ are time varying quantities and P_0 and V_0 are equilibrium values. If

$$p \ll P_0$$
$$\tau \ll V_0 \tag{19}$$

then

$$\frac{1}{P_0}\frac{\partial p}{\partial t} = \frac{-\gamma}{V_0}\frac{\partial \tau}{\partial t} \tag{20}$$

To satisfy the law of mass conservation, one writes

$$\tau = V_0 \operatorname{div} \xi \tag{21}$$

where ξ is the average vector displacement of the box.

Differentiation of Eq. (21) with respect to time and substitution of it in Eq. (20) yields

$$\frac{\partial p}{\partial t} = -\gamma P_0 \operatorname{div} \mathbf{q} \tag{22}$$

The elimination of p from Eqs. (16) and (22) yields the wave equation

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p \tag{23}$$

where, by definition

$$c^2 = \gamma P_0 / \rho_0$$

One d [illegible]

particle velocities are constant in magnitude and the particle velocities are constant in direction at all points of any given member of the family; the planes are called *wave fronts*, the direction perpendicular to them the *wave normal* If the axis of x is in the direction of the wave normal, then the wave fronts are parallel to the plane of yz.

Since p and $\mathbf{q}$ are to be constant in any one wave front, the partial derivatives with respect to y or z must vanish, and in Eq. (22):

$$\operatorname{div} \mathbf{q} = \frac{\partial q_x}{\partial x} \tag{24}$$

This equation reveals that the x component of $\mathbf{q}$ is the only component of $\mathbf{q}$ remaining in the equations basic to the wave equation so that the wave is *longitudinal.*

The one-dimensional wave equation becomes

$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{\partial^2 p}{\partial x^2} \tag{25}$$

or inertia, but rather the ability to store energy in the electrostatic and electromagnetic fields. The electrostatic field corresponds in every respect to the field of an irrotational fluid motion, and the mathematical formulation of the motions of acoustic waves and electromagnetic waves are similar.

In the customary manner, in an electromagnetic medium one defines the vectorial factors **E** as the electric field strength (analogous to potential energy) and **H** as the magnetic field strength (analogous to kinetic energy), with both magnitudes and directions at every point in space.

Next, it is assumed that the dielectric is homogeneous and isotropic (that is, that the dielectric constant κ and the permeability μ are constants), that there are no applied electromotive forces in the portion of the medium being dealt with, and that the electrical conductivity σ of the medium is zero. (Waves can be propagated in media where σ is greater than zero but less than infinity.) Then the field equations can be written

$$\frac{\kappa}{c}\frac{\partial \mathbf{E}}{\partial t} = \operatorname{curl} \mathbf{H} \qquad (1)$$

$$\frac{-\mu}{c}\frac{\partial \mathbf{H}}{\partial t} = \operatorname{curl} \mathbf{E} \qquad (2)$$

$$\operatorname{div} \mathbf{H} = 0 \qquad (3)$$

$$\operatorname{div} \mathbf{E} = 0 \qquad (4)$$

where c will be shown to be the velocity of propagation of the electromagnetic wave in vacuo, where $\kappa = \mu = 1$.

The wave equation J. C. Maxwell recognized about 1863 that these basic equations could be combined to yield an equation resembling the wave equation for mechanical wave motion Thus he predicted the existence of electromagnetic waves which had not been suspected theretofore. Later, electromagnetic waves proved to be identical with light waves.

The combination of Eqs (1) to (4) yields the wave equation

$$\frac{\kappa\mu}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = \nabla^2 \mathbf{E} \qquad (5)$$

In vacuo, the wave equation becomes

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} = c^2 \nabla^2 \mathbf{E} \qquad (6)$$

and also,

$$\frac{\partial^2 \mathbf{H}}{\partial t^2} = c^2 \nabla^2 \mathbf{H} \qquad (7)$$

where ∇^2 is a scalar operator called the Laplacian See LAPLACIAN.

Plane wave propagation. For simplicity, particular solutions of the wave equation for the case of plane wave propagation will now be sought. A wave is called plane when a family of parallel planes can be taken in the field, such that the electric and magnetic field strengths are constant in magnitude and direction at all points of any given member of the family; the planes are called wave fronts, the direction perpendicular to them the wave normal. If the axis of x is taken in the direction of the wave normal, then the wave fronts are parallel to the plane of yz

Since, for this case, **E** and **H** are to be constant in any one wave front, the partial derivatives with respect to y or z must vanish Then the field equations reveal that

$$\frac{\partial E_x}{\partial t} = \frac{\partial E_x}{\partial x} = \frac{\partial H_x}{\partial t} = \frac{\partial H_x}{\partial x} = 0 \qquad (8)$$

Therefore, insofar as wave propagation is concerned,

$$E_x = H_x = 0 \qquad (9)$$

The meaning of Eq. (9) is that neither **E** nor **H** can have a periodically changing component in the direction the wave is traveling. That is to say, the waves are not longitudinal, but rather are transverse.

Of the remaining four equations developing out of the field equations, two of them connect E_y and H_z and the other two, E_z and H_y. Dealing with the pairs, one obtains the one-dimensional wave equation

$$\frac{\partial^2(\ )}{\partial t^2} = c^2\frac{\partial^2(\ )}{\partial x^2} \qquad (10)$$

where E_y, E_z, H_y, or H_z may (any one) be inserted within the parentheses of Eq (10).

The general solution of Eq. (10) can be written in the form

$$E_y = F_1(x - ct) + F_2(x + ct) \qquad (11)$$

The first term of Eq. (11) is associated with an outward traveling wave and the second term with a backward traveling wave. This follows because the argument of F_1 will be the same whenever $x = ct$, so that as t becomes greater, the position of a wave front moves in the positive direction of x For F_2, $-x = ct$, so that the opposite is true.

If only the outward traveling wave is considered

$$E_y = F_1(x - ct) \qquad (12)$$

Such a wave is transmitted without change of shape, because its position x can always equal ct. That is, at any given elapsed time t, E_y will have the same value at $x = ct$ as it had at $x = t = 0$. Also because $c = x/t$, c has the dimensions of a velocity and is the speed at which the wave travels.

Light is also an electromagnetic process so that one should be able to state the optical properties of a substance once its electrical constants are given. However, at the high frequencies of visible light, κ and μ vary with frequency and this variation must be taken into account if Maxwell's equations are to give a description of optical phenomena which fits measured data

Acoustic waves in gases. In gases, the existence of inertia (resulting in kinetic energy) and elasticity (resulting in potential energy) are obvious For example, imagine tak[illegible] e medium a small packet of gas encl[illegible] less

and in the direction of its longitudinal axis. The wave equation is identical to Eq. (29) except that ξ_t replaces ξ_v. The motions of the particles of the bar are in a line with the longitudinal axis of the bar. Longitudinal and transverse waves may, and generally do, coexist.

Waves in plates Wave motion in a plate is similar to that in a bar. The bending of a plate compresses the material on the inside of the bend and stretches it on the outside. But when a material is compressed it tries to spread out in a direction perpendicular to the compressional force, so that when a plate is bent downward in one direction there is a tendency for it to curl up in a direction at right angles to the bend. The ratio of the sidewise spreading to the compression is called Poisson's ratio and is designated by the letter σ.

The wave equation for the plate is given by

$$\nabla^4\eta + \frac{12}{(hc_L)^2}\frac{\partial^2\eta}{\partial t^2} = 0 \qquad (32)$$

where η is the displacement of the plate perpendicular to its surface, h is the thickness of the plate, and c_L is the longitudinal plate velocity

$$c_L = \sqrt{\frac{Y}{\rho(1-\sigma^2)}} \qquad (33)$$

From a solution to the equation it is possible to find the velocity of propagation for the transverse (bending) wave.

$$c_B = \sqrt{1.8hc_L}\sqrt{f} \qquad (34)$$

Just as for the bar, the velocity of propagation is dependent on frequency, and the plate is said to be a dispersing medium.

Steady state wave motion. Very frequently the source of a wave is steady and the wave produced by it is periodic, at least for a long period of time compared to the build-up and decay times of the wave. When this is true, it is convenient to specify the functions $F_1(x-ct)$ and $F_2(x+ct)$ of Eq. (11) each by a summation of sinusoidal functions:

$$F(x-ct) = \sum_{\nu} A_\nu \cos\left[\omega_\nu\left(t-\frac{x}{c}\right)+\theta_\nu\right] \qquad (35)$$

where $\omega_\nu = 2\pi f_\nu$; f_ν is the frequency of the νth component of the wave, and θ_ν is the phase angle of the νth component, which is determined when $x = t = 0$.

From the well known theory of Fourier series, any function $F_1(x-ct)$, if it repeats on itself periodically, can be represented by a linear summation of sine or cosine wave functions (*see* FOURIER SERIES AND INTEGRALS). The component with the lowest frequency f_1 is the fundamental component or the first harmonic; each higher component, called the $(\nu-1)$th overtone, or the νth harmonic, has a frequency equal to νf_1. In other words, ν is an integer and the wave form $F_1(x-ct)$ is represented by a linear summation of cosine terms with frequency components f, that are harmonically related (f, $2f$, $3f$, $4f$, and so forth) and at $t = x = 0$ have phase angles θ_1, θ_2, θ_3, and so forth, that may differ from zero.

As a further simplification, each component of Eq. (35) is usually written

$$P = A\cos(\omega t - kx + \theta) \qquad (36)$$

where $k = \omega/c$ is called the wave number. The period of this component is T and equals $1/f$. As before, θ is the phase angle in radians.

Note that A is the peak amplitude of the component wave. Generally, in acoustics, measuring devices read the root mean-square value of a wave, so that the intensity of the wave is designated by $A_{rms} = A/\sqrt{2}$, and the strength of a wave represented by Eq. (35) is given by

$$A_{rms} = \sqrt{A^2_{1_{rms}} + A^2_{2_{rms}} + \cdots} \qquad (37)$$

See ROOT-MEAN-SQUARE [L.L.B.]

Bibliography: M. Abraham and R. Becker, *Classical Theory of Electricity and Magnetism*, 1937; L. L. Beranek, *Acoustics*, 1954; P. M. Morse, *Vibration and Sound*, 2d ed., 1948; Lord Rayleigh, *The Theory of Sound*, 2d ed. (reprint), 1945; J. C. Slater and N. H. Frank, *Electromagnetism*, 1947.

Wave motion in fluids

Wave motion is the basic mechanism by which local disturbances are propagated from one part of a [illegible]

without net mass motion. The direction along which the local disturbances are transmitted is called the direction of propagation, and the speed of the disturbances relative to the fluid is called the wave speed, or the speed of propagation.

Applications. Wave phenomena have widespread applications. Marine equipment such as the Fathometer and Sofar rely upon wave propagation (*see* ECHO SOUNDER). Shock waves used in Sofar propagate long distances under water, being refracted by the isothermal layers in the oceans (*see* SOFAR). The study of waves is directly applicable to supersonic aircraft, wind tunnels, shock tubes, rocket combustion oscillation, nuclear bomb blasts, controlled fusion processes in plasma, and ultrasonic processes such as cleaning and inspection.

Waves in fluids are diffracted and refracted so that they can be focused to produce intense concentration of energy. For example, in shock tubes, energy is focused into a sharp pulse, the fluid there being caused to glow by the intense excitation so produced. Ultrasonic waves are focused and directed through tanks of mercury to provide short-term memory for large-scale computers.

Fluid waves caused by successive firing of charges from a vertical sounding rocket, such as the Aerobee as it climbs above the altitude in which ballons are effective, are recorded on the ground to provide data for the determination of wind velocity and air density profiles in the upper atmosphere. The technique is analogous to seismic

The general solution to Eq. (25) is exactly the same as that for electromagnetic waves and is given by Eq. (11). The discussion following Eq (11) is also valid here.

One-dimensional spherical waves. In free space, it is frequently desired to express mathematically the radiation of sound from a nondirectional source. In this case the sound wave expands as it travels away from the source and the wave front is always a spherical surface. The operator on the right side of Eq. (23) can be written in a form suitable to spherical coordinates.

Assuming equal radiation in all directions, the wave equation in one-dimensional spherical coordinates then becomes

$$\frac{\partial^2 p}{\partial r^2} + \frac{2}{r}\frac{\partial p}{\partial r} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{26}$$

Differentiation shows that Eq. (26) can also be written

$$\frac{\partial^2(pr)}{\partial t^2} = c^2\frac{\partial^2(pr)}{\partial r^2} \tag{27}$$

Equation (27) has the same form as Eq. (25) Hence, the same formal solution will apply to either equation except that the dependent variable is $p(x,t)$ in one case and $pr(r,t)$ in the other case

The solution to Eq (27) for the outward traveling wave only (free space) is, then

$$p = \frac{1}{r}F_1(r - ct) \tag{28}$$

Note that, just as for the plane wave, the wave is propagated without change of shape However, the magnitude of the sound pressure decreases inversely with distance due to the spreading of the wave as it propagates

Acoustic waves in liquids. Acoustic waves in liquids obey the same equations as those in gases The velocity of propagation is greater in liquids than in gases, and viscous losses are often higher. *For information on an important example of* wave motion in liquids, *see* UNDERWATER SOUND, *see also* WAVE MOTION IN FLUIDS; WAVE MOTION IN LIQUIDS.

Acoustic waves in solids. Several different types of acoustic waves may exist in solids, depending on the different manners in which potential energy is stored in the solid The wave equations associated with several of these types are reviewed in the following paragraphs.

Waves on flexible stretched strings. The wave equation for a flexible stretched string is given by

$$\frac{\partial^2 \xi_y}{\partial t^2} = c^2\frac{\partial^2 \xi_y}{\partial x^2} \tag{29}$$

where ξ_y is the displacement of the string at a point x along the string in the y direction (perpendicular to the string). The speed of propagation c is equal to the square root of the ratio of the tension (in dynes) to the linear density of the string (in g/cm²).

The solution to this equation is identical to that for Eq. (10). Because the motion of the elements of the string is perpendicular to the string, at least for small displacements, the waves are said to be transverse.

Acoustic waves in bars. For the purposes of this article, it is assumed that a string has tension with negligible stiffness, while a bar has stiffness without tension. When a bar is bent, its lower half is compressed and its upper half is stretched, or vice versa. When the bending force is removed, the bar attempts to regain its equilibrium position. The restoring force is due to the moment of the forces about the neutral plane in the bar and is related to the cross-sectional dimensions and the Young's modulus of the material.

The transverse wave equation describing the motion of such a bar is

$$\frac{\partial^2 \xi_y}{\partial t^2} = -\kappa^2\frac{Y}{\rho}\frac{\partial^4 \xi_y}{\partial x^4} \tag{30}$$

Where ξ_y is the displacement perpendicular to the neutral plane of the bar, Y is Young's modulus, ρ is the density of the bar, and κ is the radius of gyration of the cross section. Values of κ for some of the simpler cross-sectional shapes are as follows:

Rectangle: Length b parallel to center line, width a perpendicular to center line,

$$\kappa = a/\sqrt{12}$$

Circle: Radius a, $\kappa = a/2$

Circular ring: Outer radius a, inner radius b,

$$\kappa = 0.5(a^2 + b^2)^{1/2}$$

Equation (30) differs from the usual wave equation, Eq. (10), in that it has a fourth derivative with respect to x instead of a second derivative.

The function $F_1(x - ct)$ is not a solution, so that a bar, satisfying Eq (30), cannot have waves traveling along it with a velocity independent of frequency and with an unchanged shape.

If one considers excitation of the bar by one frequency f at a time, a solution to Eq. (30) may be written as follows:

$$\xi_y = A\cos[2\pi(\mu x - ft) + \phi] \tag{31}$$

where $\mu = (f^2\rho/4\pi^2 Y\kappa^2)^{1/4}$, $f = 2\pi\mu^2\kappa\sqrt{Y/\rho}$, A = amplitude, and ϕ = phase angle.

The velocity of propagation of the wave is equal to $(f/\mu) = (4\pi^2 Y\kappa^2/\rho)^{1/4}\sqrt{f}$. It obviously depends on the frequency f of the exciting wave Such a velocity for a simple harmonic wave is called the phase velocity (*see* PHASE VELOCITY). For a complex wave with several components differing in frequency, the shape must change with distance of travel. A bar is sometimes said to be a dispersing medium for waves of bending.

There is also a possibility for a longitudinal wave in such a bar. In this case t   is excited at an

The above result has been found to agree so well with experimental observations under ordinary conditions that measurement of the speed of sound has become one of the standard methods for determining the value of γ for various gases

For liquids, and for gases at extreme temperatures and densities, or for disturbances of very high frequencies, the adiabatic law loses its usual significance, so that the speed of sound (or the compressibility $d\rho/dp$) has to be obtained either from direct measurement or from more exact theories.

The speed of sound in several common fluids at room temperature is given in the accompanying list together with the speed of sound in some common solids for comparison.

Substance	*Speed of sound, ft/sec*
Air	1,130
Hydrogen	4,320
Water	4,800
Mercury	4,600
Paraffin	4,300
Lead	4,030
Aluminum	16,700
Iron	16,800

At this point, it may be well to give some numerical examples on sound amplitude. The smallest periodic pressure amplitude detectable by the human ear is in the order of 10^{-3} dyne/cm², or 10^{-9} atm (at about 2000 cps). On the other hand, the threshold of feeling corresponds to a pressure amplitude roughly 10^6 times greater Thus, these limits of what is ordinarily called sound corresponded to pressure fluctuations of 10^{-9} to 10^{-3} atm, which justified the assumption of small disturbance in the foregoing derivation of the speed of sound.

Zone of action and zone of silence. Small disturbances can only propagate in a fluid at a finite speed. The same is true for disturbances of large amplitude (see SHOCK WAVE). Therefore, when an object moves through a stationary fluid at a speed in excess of the speed at which disturbances can be propagated, there will be a boundary that divides the fluid into two distinct regions: namely, the zone of action that has, and the zone of silence that has not yet, been affected by the motion of the object at any given instant To understand this behavior, consider Fig. 2.

Let u denote the speed of the object and a denote the speed of propagation of the fluid disturbance To simplify the discussion, assume that the object is in uniform rectilinear motion, and the amplitude of the disturbance is so weak that a can be identified with the speed of sound in the stationary fluid At any instant t, when the object is at a certain point P, the object will generate disturbances which will propagate away from P in all directions with velocity a After time interval Δt, the object would have traveled a distance $u\Delta t$ and would have moved to new position P' along its trajectory, while the disturbances generated at point P would still be confined to a spherical surface of radius

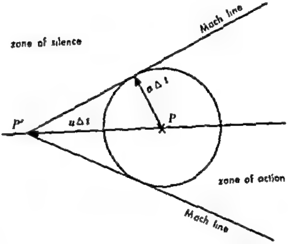


Fig 2 A point source of disturbance in uniform rectilinear motion

$a\Delta t$ Because u is greater than a, all disturbances generated by the object up to the time $t + \Delta t$ will be confined within the conical surface of half vertex angle

$$\beta = \sin^{-1} (a/u) \tag{8}$$

with vertex at point P' and axis along PP'. The region inside such a conical surface is called the

of silence at the same instant

The ratio between the speed of the object and the speed of sound u/a is known as the Mach number of the moving object, and is usually denoted by the symbol M The motion of the object is accordingly called subsonic when $M < 1$; transonic when $M \cong 1$, supersonic when $M > 1$; and hypersonic when $M \gg 1$ From Eq. (8) and from Fig. 2, which has been depicted for supersonic motion, it can be deduced that the zone of action extends to all parts of the fluid for subsonic motion. For transonic motion, the zone of action covers approximately the rear half of the fluid, up to the plane tangent to the nose of the moving object; for hypersonic motion, the zone of action is confined to a relatively slender cone about the trajectory of the object.

Waves of larger amplitude. In physical optics, it is well known that wavelets from a distributed light source in space can be superposed according to Huygens' principle. For wave motion in fluids, however, the analogy holds only as long as the amplitude of the resultant disturbance remains small enough for the acoustic approximation to apply. For waves of larger amplitude, the nonlinear behavior of the fluid dynamic equations must be taken into account To visualize these nonlinear effects, consider the following one dimensional problem

A semi-infinite tube is filled with a gas initially at rest At one end of this tube is a movable piston. The piston is suddenly given a small velocity toward the gas. As a result, a compression wave is

exploration for petroleum. Waves in the upper atmosphere are being found to influence weather and their recent Fourier analysis by meteorologists has greatly advanced the techniques of long-range weather forecasting. Shock waves are used to prepare free radicals and in forcing certain chemical processes

Wave classification. In contrast to surface waves (*see* OCEAN WAVES), which are transverse oscillations caused by gravity acting on a liquid having a free surface, wave motion within a fluid is generated by successive compression and expansion of adjacent volume elements of a compressible fluid. Because compression and expansion of an ordinary fluid can only proceed along the direction of propagation of the disturbance, waves within a fluid are mostly longitudinal waves.

Waves in a fluid can be classified as compression waves and expansion waves, according to whether the disturbance is a compression or an expansion. They can further be classified according to the amplitude of the disturbance and the chemical nature of the fluid. For example, waves of small amplitude are called acoustic (or sound) waves; compression waves propagating in chemically inert fluids are called shock waves; waves propagating in the earth are seismic waves; waves of large amplitude generated by rapid chemical reactions in explosive fluids are called detonation waves; they can propagate much faster than sound waves Waves in an electrically conducting fluid in the presence of strong magnetic fields are called magnetohydrodynamic waves.

Acoustic waves. The acoustical wave equation can be derived formally through linearization of the equations of motion A more straightforward approach is to consider a plane wavefront that moves from right to left at a constant speed u_1 in a fluid which is initially at rest and of density ρ_1. If an observer fixes his attention on the wavefront by also moving from right to left at the same speed u_1, he will witness a steady flow of fluid from left to right across the wavefront (Fig. 1) Represent the flow velocity and the fluid density to the right of the wavefront by u_2 and ρ_2, respectively; then conservation of mass requires that

$$\rho_1 u_1 = \rho_2 u_2 \tag{1}$$

because no fluid mass can accumulate at the wavefront in a steady state. Furthermore, an increase in fluid momentum across the wavefront can only be supported by a corresponding drop in pressure from p_1 to p_2 in the fluid. Therefore

$$\rho_2 u_2^2 - \rho_1 u_1^2 = p_1 - p_2 \tag{2}$$

If the disturbance is so weak that the fractional changes in flow velocity, fluid density, and pressure across the wavefront are much smaller than unity, the changes can be written as

$$u_2 = u_1 + du$$
$$\rho_2 = \rho_1 + d\rho$$
$$p_2 = p_1 + dp$$

By substituting these into Eqs (1) and (2) and neglecting the product terms of the differential quantities, Eqs. (1) and (2) become

$$\rho_1 du + u_1 d\rho = 0 \tag{1a}$$
$$2\rho_1 u_1 du + u_1^2 d\rho = -dp \tag{2a}$$

An expression for u_1 is obtained by eliminating du from these two equations.

$$u_1^2 = dp/d\rho \tag{3}$$

Because this derivation began with a wavefront moving in a fluid initially at rest, the above result shows that any small disturbance, if propagated by wave motion at all, must propagate relative to the fluid at the speed of sound a where

$$a = u_1 = \sqrt{dp/d\rho} \tag{4}$$

If the disturbance is periodic in time with a fundamental frequency ν, as in a musical note, then there will be a wavelength λ associated with the wave motion, $\lambda = a/\nu$. Furthermore, Eq (4) shows that a depends only on the variation of pressure with density as caused by a small mechanical disturbance in the fluid, and that a is real as long as $dp/d\rho$ is positive Historically, Isaac Newton first attempted to derive the speed of sound in air from a somewhat different approach. He arrived at a result which was equivalent to assuming an isothermal compression process Thus, from the equation of state for a perfect gas (Boyle's law),

$$p = \rho RT \tag{5}$$

he obtained

$$a = \sqrt{RT} = \sqrt{p/\rho} \tag{6}$$

However, experimental measurements of the speed of sound in air turned out to be consistently higher than his prediction This led P. S. Laplace to suspect that compression and expansion processes associated with acoustic waves should obey the adiabatic law $p\rho^{-\gamma}$ = constant ($\gamma = C_p/C_v$, being the ratio of specific heats), instead. If this pressure-density relationship is assumed, the speed of sound then becomes

$$a = \sqrt{\gamma RT} \tag{7}$$

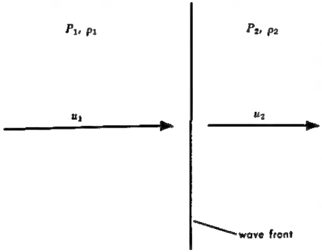


Fig 1. Flow across a plane wavefront.

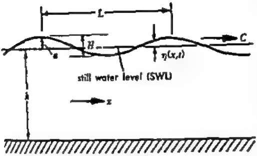


Fig 1 Definition sketch for an oscillatory wave.

length will be a function of the period of oscillation The two major classes of oscillatory waves, deep-water and shallow-water waves, are determined by the magnitude of the ratio of liquid depth to wavelength h/L. An inspection of the celerity equation for gravity waves shows that as the depth becomes large in comparison with wavelength,

$$\tanh 2\pi\frac{h}{L} \to 1 \quad \text{and} \quad C = \sqrt{\frac{gL}{2\pi}}$$

hence, in a deep-water wave the celerity is a function only of the wavelength. This approximation is close if the depth is greater than one-half the wavelength On the other hand, as the depth becomes small in comparison with wavelength,

$$\tanh 2\pi\frac{h}{L} \to 2\pi\frac{h}{L} \quad \text{and} \quad C = \sqrt{gh}$$

hence, the celerity depends only on the depth in a shallow-water wave. Somewhat arbitrarily, the limit $h/L = 1/10$ is generally applied to this type of wave motion In deep-water waves, individual fluid particles tend to move in circular orbits The radius of the surface particle orbit is equal to the wave amplitude and the radius decreases exponentially with depth (Fig 2). At a depth of one-half the wavelength the orbital radius is about $1/20$ of the amplitude A zone of essentially zero fluid motion is rapidly approached and the character of the wave is therefore not affected by the total depth of the liquid.

In shallow water, no vertical particle motion can exist at the bottom; thus the wave characteristics are modified The particle orbits are flat ellipses in which the minor axis is depressed to zero at the bottom (Fig 2).

The energy of a wave consists of equal amounts of potential energy (due to particle position above or below the still water level) and kinetic energy (due to the motion of particles in their orbits). The rate of propagation of energy in the direction of wave travel is known as the group velocity to distinguish it from phase velocity C. In deep-water waves the group velocity is one-half the phase velocity; in shallow-water waves the two propagation velocities are equal

Standing waves. A standing wave can be considered to be composed of two equal oscillatory wave trains traveling in opposite directions. The phase velocity of the resulting wave is zero; nevertheless the velocity of propagation of the component waves retains its usual meaning In the notation of the previous section, the equation for the profile of a standing wave is obtained by adding the elevations of waves moving in the positive and negative x directions respectively

$$\vec{\eta}_1 = a \sin 2\pi\left(\frac{t}{T} - \frac{x}{L}\right)$$

and

$$\overleftarrow{\eta}_2 = a \sin 2\pi\left(\frac{t}{T} + \frac{x}{L}\right)$$

hence

$$\eta = \eta_1 + \eta_2 = H \sin 2\pi\frac{t}{T} \cos 2\pi\frac{x}{L}$$

If the length of the basin l in which a disturbance occurs is an integral number n of half wavelengths, a self-perpetuating (except for frictional dissipation) standing wave will result. Therefore, if $l = nL/2$

$$\eta = H \sin 2\pi\frac{t}{T} \cos\frac{\pi n x}{l}$$

For long waves as with shallow water, in a basin of uniform depth, the period of oscillation T is defined by

$$T = 2l/n\sqrt{gh}$$

Standing waves frequently occur in canal locks as a result of filling disturbances and in large lakes, bays, and estuaries as a result of wind or tidal action *See* SEICHE.

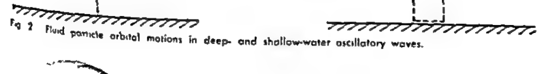


Fig 2 Fluid particle orbital motions in deep- and shallow-water oscillatory waves.

generated and propagates along the tube at the speed of sound of the undisturbed gas. Suppose that, soon after this wave has started, the piston is given an additional increment in velocity. A second wave is formed and propagates along the tube behind the first. Because the pressure has increased slightly across the first wave, so has the temperature and the speed of sound. Furthermore, the gas behind the first wave is already moving along the tube at the piston velocity after the first impulse. Thus, the second wave, which propagates with respect to a moving gas ahead of itself and at a sound speed slightly higher than that of the first wave, will soon catch up with the first wave.

If the accelerating motion of the piston continues, the succeeding wavelets will propagate along the tube at increasingly higher velocity, so that the shape of the resultant compression pulse from the piston motion will appear steeper as it progresses. Similarly, a decelerating piston motion produces an expansion pulse that flattens out as it progresses along the tube It is this asymmetry between the two processes that makes the occurrence of large amplitude compression waves, called shock waves, a more noticeable phenomenon in nature than expansion waves.

Shock waves are characterized by rapid changes in fluid density, pressure, and temperature along the direction of flow Bomb blasts start as shock waves

Seismic waves. Seismic waves pass through the ground; they arise either from natural readjustment of the faults in the earth's crust or by man-made explosions. According to their modes of propagation, seismic waves may be divided into two main groups, namely body and surface waves.

Body waves, which propagate through the inside of the earth, may further be subdivided into dilation (longitudinal) waves, which are similar to acoustic waves in compressible fluids, and shear (transverse) waves, which arise on account of the large shear resistance of most elastic solids For any given medium, dilation waves usually have approximately twice the velocity of propagation of shear waves.

Surface waves from any distant earthquake always arrive after both the dilation waves and the shear waves, and they normally register a much larger amplitude signal on the seismogram. From the known relationship between the propagation velocities and the mechanical properties of various substances, seismologists extract valuable information about the structure of the earth's interior from the seismograms and apply the results to such purposes as mine prospecting. *See* SEISMOLOGY; SONIC BARRIER; SONIC BOOM. [S.C.L.]

Bibliography: R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves*, 1948; S. Feldman, *Hypersonic Gas Dynamic Charts for Equilibrium Air*, Avco Research Laboratory, 1957; A. R. Kantrowitz (ed.), *High Speed Aerodynamics and Jet Propulsion*, vol. 3, 1958; R. J. S. Rayleigh, *Theory of Sound*, 2d rev. ed., 1945.

Wave motion in liquids

A temporal variation in fluid velocity which is propagated through a fluid medium The speed of propagation of the disturbance, or change in fluid velocity relative to the initial velocity of the fluid medium, is known as the wave celerity. The dynamic behavior of the wave depends upon the method of generation of the disturbance, the boundary conditions of the fluid medium, and the fluid properties. This article treats those aspects of wave motion appropriate to liquids in which disturbances are propagated at a gas-liquid interface and are primarily dependent upon the gravitational fluid property (surface tension and viscosity being of secondary importance). Wave motions which occur in confined fluids (either liquid or gaseous) are primarily dependent upon the elastic property of the medium. *See* WAVE MOTION IN FLUIDS; WATER HAMMER

The fundamental concepts of gravity waves in liquids are presented in the one-dimensional form For information on the generation of waves by wind in natural bodies of water and their transformation in coastal regions, *see* OCEAN WAVES; SEA STATE; SHORE PROCESSES; WAVE (CAPILLARY) WAVE (INTERNAL)

Oscillatory waves. The term oscillatory implies a periodicity in the form of a disturbance moving past a fixed point Figure 1 is a definition sketch for an oscillatory wave propagating in a liquid of constant density ρ and depth h measured from the bottom to the still water level (SWL). Wave length L is the horizontal distance between successive crests of the wave Wave height H is the vertical distance from crest to trough; amplitude is the distance from the still water level to the crest, and η is the elevation of the free surface with respect to the still water level at any position x and instant of time t. In the linearized theory of small-amplitude waves ($H/L < .03$) the wave profile is sinusoidal and is given by

$$\eta = a \sin 2\pi \left(\frac{t}{T} - \frac{x}{L}\right)$$

where T is the wave period By definition, the celerity or speed of propagation $C = L/T$ and is given by

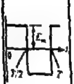
$$C = \sqrt{\left(\frac{\sigma}{\rho}\frac{2\pi}{L} + \frac{gL}{2\pi}\right) \tanh 2\pi \frac{h}{L}}$$

The first term on the right expresses the influence of surface tension σ and need be considered only for waves of very small length (of the order of magnitude of 1 in) In the remaining development only gravity waves will be considered, as expressed by the second term of the celerity equation.

Oscillatory waves may be generated in a rectangular channel by a simple harmonic translation of a vertical wall forming one end of the flume The wave amplitude will be determined by the displacement (stroke) of t -and the wave

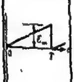
$E_m \sin(2\pi t/T)$

Sine wave



$$\frac{4E_m}{\pi}\sum^{\infty}\frac{1}{n}\sin(2\pi nt/T)$$
$n = 1, 3, 5, \ldots$

Square wave



$$\frac{E_m}{2}-\frac{E_m}{\pi}\sum^{\infty}\frac{1}{n}\sin(2\pi nt/T)$$
$n = 1, 2, 3, \ldots$

Saw-tooth wave

$$\frac{8E_m}{\pi^2}\sum^{\infty}(-1)^{(n-1)/2}\frac{1}{n^2}\sin(2\pi nt/T)$$
$n = 1, 3, 5, \ldots$

Triangular wave

$$\frac{E_m}{\pi}\left[1+\frac{\pi}{2}\sin(2\pi t/T) - 2\sum^{\infty}\frac{1}{4n^2-1}\cos(4\pi nt/T)\right]$$
$n = 1, 2, 3, \ldots$

Half-wave rectified sine wave

$$\frac{2E_m}{\pi}\left[1-2\sum^{\infty}\frac{1}{4n^2-1}\cos(2\pi nt/T)\right]$$
$n = 1, 2, 3, \ldots$

Full-wave rectified sine wave

frequency. The Fourier series for each waveform is given beside each figure as functions of time t, where E_m is the maximum value of the wave and T is the period. *See* WAVEFORM, NONSINUSOIDAL.

Sine waves are obtained from sine-wave generators and *LC*, *RC* and beat-frequency oscillators. *See* ALTERNATING-CURRENT GENERATOR; OSCILLATOR; SINE WAVE.

Square waves are obtained from square-wave generator circuits, such as multivibrators *See* MULTIVIBRATOR; SQUARE WAVE.

Saw-tooth waves are obtained from gas-tube relaxation oscillators, and thyratron and vacuum-tube sweep circuits. *See* RELAXATION OSCILLATOR; SAW-TOOTH WAVE, SWEEP GENERATOR.

Triangular waves are obtained from integrated square waves.

The output wave shape of a half-wave rectifier with resistance load is as illustrated. *See* RECTIFIER.

The output wave shape of a full-wave rectifier with resistance load is as illustrated. [D.L.W.]

Waveform, nonsinusoidal

Electric circuits containing nonlinear elements, such as electron tubes, iron-core magnetic devices, and transistors, commonly produce nonsinusoidal currents and voltages When these are repetitive functions of time, they are called nonsinusoidal electric waves. Oscillograms, tabulated data, and sometimes mathematical functions for segments of such waves are often used to describe the excursions throughout one cycle. A cycle corresponds to 2π electrical radians and covers the time interval T seconds (sec) in which the wave repeats itself

These electric waves can be represented by a constant term, the average or dc component, plus a series of harmonic terms in which the frequencies of the harmonics are integral multiples of the fundamental frequency The fundamental frequency, f_1, if it does exist, has the time span of $T = 1/f_1$ sec for its cycle. The second-harmonic frequency, f_2, then will have two of its cycles within T sec, etc.

Fourier series representation. The series of terms stated above is known as a Fourier series and can be expressed in the form.

$$\begin{aligned} y(t) &= B_0 + A_1 \sin \omega t + A_2 \sin 2\omega t + \\ &\quad + A_n \sin n\omega t + B_1 \cos \omega t + B_2 \cos 2\omega t \\ &\quad + \cdots + B_n \cos n\omega t \\ &= B_0 + C_1 \sin(\omega t + \phi_1) + \cdots \\ &\quad + C_n \sin(n\omega t + \phi_n) \\ &= \sum_{n=0}^{\infty} C_n \sin(n\omega t + \phi_n) \end{aligned} \tag{1}$$

where $y(t)$, plotted over a cycle of the fundamental, gives the shape of the nonsinusoidal wave. The terms on the right-hand side show the Fourier series representation of the wave where

$$C_n = \sqrt{A_n^2 + B_n^2} \qquad \text{and} \qquad \phi_n = \arctan \frac{B_n}{A_n}$$

A_0 is identically zero, and $B_0 = C_0$ in Eq (1).

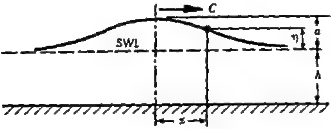


Fig. 3. Definition sketch for a solitary wave

Solitary waves. A solitary wave consists of a single crest above the original liquid surface which is neither preceded nor followed by another elevation or depression of the surface. Such a wave is generated by the translation of a vertical wall starting from an initial position at rest and coming to rest again some distance downstream. In practice, solitary waves are generated by the motion of barges in narrow waterways or by a sudden change in the rate of inflow into a river; they are therefore related to a form of flood wave. The amplitude of the wave is not necessarily small compared to the depth, and the wavelength is theoretically infinite because the elevation of the surface approaches the still water level asymptotically with distance as shown in Fig. 3. The profile of the solitary wave is given by

$$\eta = a \operatorname{sech}^2 \left[\frac{x}{h}\sqrt{\frac{3}{4}\frac{a}{h}}\right]$$

and the celerity by

$$C = \sqrt{g(h+a)}$$

When the solitary wave amplitude becomes approximately equal to the depth, the wave profile becomes unstable and a breaking wave results.

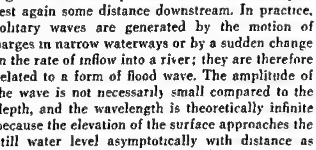


Fig. 4. Definition sketch for a surge wave.

Surges. A surge is generated by the forward motion of a vertical wall, at a constant speed, as shown schematically in Fig. 4. Surges in open channels are analogous to shock waves produced in a tube by the continuous motion of a piston. A zone of violent eddy motion occurs at the wavefront and the analysis of such motions must take into account the appreciable energy dissipation in this region. The velocity of propagation of a surge is given by

$$C = \sqrt{gh_1}\left[\frac{1}{2}\frac{h_2}{h_1}\left(\frac{h_2}{h_1}+1\right)\right]^{1/2}$$

If a velocity V_1 equal and opposite to C is imposed on the fluid upstream of the disturbance, the absolute velocity of the surge front will become zero. In this form the surge is known as a hydraulic jump and it is frequently employed as a means of dissipating flow energy at the bottom of dam spillways. See HYDRAULIC JUMP. [D.R.F.H.]

Bibliography: R. C. H. Russell and D. H. Macmillan, *Waves and Tides*, 1953; J. J. Stoker, *Water Waves*, 1957.

Wave optics

The branch of optics which treats of light (or electromagnetic radiation in general) with explicit recognition of its wave nature. The counterpart to wave optics is ray optics or geometrical optics, which does not assume any wave character but treats the propagation of light as a straight-line phenomenon except for changes of direction induced by reflection or refraction. See OPTICS; OPTICS, GEOMETRICAL.

Any optical phenomenon which is correctly describable in terms of geometrical optics can also be correctly described in terms of wave optics. However, the many phenomena of interference, diffraction, and polarization are incontrovertible evidence of the wave nature of light, and geometrical optics often gives an incomplete or incorrect description of the behavior of light in an optical system. This is especially true if changes of refractive index occur within a space which is of the order of several wavelengths of the light. See DIFFRACTION; INTERFERENCE OF WAVES; POLARIZED LIGHT. [R.C.L.]

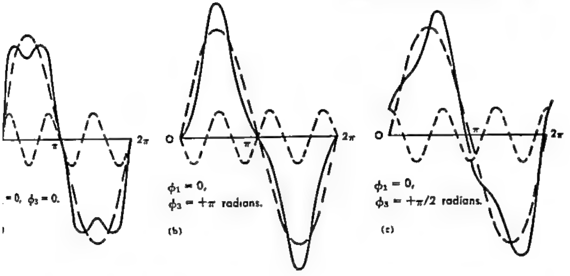
Wave packet

In wave phenomena, a superposition of waves of differing wavelengths, so phased that the resultant amplitude is negligibly small except in a limited portion of space whose dimensions are the dimensions of the packet. If the reciprocal wavelengths $k = \lambda^{-1}$ forming a one-dimensional packet lie in a band Δk, the minimum dimension of the packet is $\Delta x \cong (2\pi\Delta k)^{-1}$. When all the component waves move in the same direction, the packet speed is the group velocity $v_g = df/dk$ evaluated at the mean k, f is the frequency. When the phase velocity c depends on λ, $v_g \neq c$, and Δx changes with time. See GROUP VELOCITY; PHASE VELOCITY; QUANTUM MECHANICS; QUANTUM THEORY, NONRELATIVISTIC [E.C.]

Waveform

The pictorial representation of the form or shape of a wave, obtained by plotting the amplitude of the wave with respect to time. There are an infinite number of possible waveforms. Some of the more common electrical waveforms are illustrated here. These diagrams are plots of voltage against time. It is equally possible to show current waveforms.

It is possible to represent any periodic waveform mathematically as a Fourier series of sine and cosine terms at harmonic frequencies. Any nonsinusoidal wave is composed of a constant or dc term plus a series of harmonic terms in which the frequencies are integral m[illegible] ndamental



2 (a-c) Addition of a fundamental and a ...rd harmonic

...ave symmetry is absent, quarter-wave symmetry ...ay exist. A wave which is the same from π to ... as it is from 0 to π, that is, $y(\theta) = y(\theta + \pi)$, ...unsymmetrical. Only even harmonics are present, ...d the wave has no fundamental component. The ...utput from a full wave rectifier, for example, ...ontains only the average term, B_0, and even har...monics.

Waves that do not meet any of the special condi...tions noted above can be expected to contain both ...even and odd harmonics, and probably both sines ...nd cosines also. Any doubt arising on the har...monic content of a wave is resolved by assuming ...ll components of the Fourier series to be present ...nalysis will show exactly those that actually exist

Even and odd functions. The time origin of a ...wave can be chosen arbitrarily. If the reference ...axis, $t = 0$, is such that the wave to the left is ...merely the wave to the right rotated about this ...axis, then $y(-\theta) = y(\theta)$, which is said to be an ...even function Only cosine terms will be present ...in the Fourier series for the wave On the other hand, if the wave to the left is the wave to the right ... about ... is sai... appea... ...eries. Neither case precludes the possibility of the presence of both even and odd harmonics.

RMS value of a nonsinusoidal wave. A nonsinusoidal wave has an rms value obtained through the following steps

1 Combine all terms of the same frequency so as to have a single $A_1, A_2, \ldots, A_n$; $B_1, B_2, \ldots$, B_n, or a single $C_1, C_2, \ldots, C_n$ Terms such as $\sin(n\omega t \pm \alpha)$ and $\cos(n\omega t \pm \beta)$ each have sine and cosine components which can be separated out by trigonometric expansion.

2. Form the series $y(\theta)$ as in Eq. (1).

3. The rms value of the wave then is

If y_{rms} represents a voltage or a current, this value is shown by an electrodynamometer or iron-vane voltmeter or ammeter The rms of the wave is merely the square root of the sum of the squares of the rms values of all of the frequency components

Power. An indicating wattmeter with a nonsinusoidal voltage impressed on its potential circuit and a nonsinusoidal current in its current coils will indicate the average power taken by the circuit. Designating peak values of the component voltages and currents by E_n's and I_n's in place of C_n's,

$$\text{Average power} = \frac{1}{2\pi}\int_0^{2\pi} ei\,d\theta$$
$$= E_0 I_0 + \frac{E_1 I_1}{2}\cos\theta_1 + \cdots + \frac{E_n I_n}{2}\cos\theta_n$$

Each coefficient is simply the product of rms voltage and current No cross-product terms involving different frequencies result from the integration. That is, no power can be contributed by a voltage of one frequency and a current of another frequency

Power factor. The apparent power taken by a circuit carrying nonsinusoidal voltage and current is the product of the rms values of these quantities. Power factor for such a case is defined only by the ratio of the average power to the apparent power Thus,

$$\text{Power factor (pf)} = \frac{\text{watts average power}}{\text{rms volts} \times \text{rms amperes}}$$

Power factor hence is the ratio of instrument readings as stated All circuits have a power factor, but pf = $\cos\theta$ only for a sine wave voltage and current of the same frequency There is no average or representative phase angle for a circuit carrying nonsinusoidal waves.

$$y_{rms} = \sqrt{B_0^2 + \frac{A_1^2}{2} + \frac{A_2^2}{2} + \cdots + \frac{A_n^2}{2} + \frac{B_1^2}{2} + \frac{B_2^2}{2} + \cdots + \frac{B_n^2}{2}} = \sqrt{B_0^2 + \frac{C_1^2}{2} + \frac{C_2^2}{2} + \cdots + \frac{C_n^2}{2}}$$

The radian frequency of the fundamental is $\omega = 2\pi f_1$, and n is either zero or an integer. C_1 is the amplitude of the fundamental ($n = 1$), and succeeding C_n's are the amplitudes of the respective harmonics having frequencies corresponding to $n = 2, 3, 4$, etc., with respect to the fundamental. ϕ_1 is the phase angle of the fundamental with respect to a chosen time reference axis, and the succeeding ϕ_n's are the phase angles of the respective harmonics.

The equation for $y(t)$ shows all its separate components, which, in general, includes an infinite number of terms. In order to represent a given nonsinusoidal wave by a Fourier series, it is necessary to evaluate each term, that is, B_0, and all A_n's and all B_n's. In practical problems the first several terms will usually yield an approximate result sufficiently accurate for portrayal of the actual wave The degree of accuracy desired in representing faithfully the actual wave determines the number of terms that must be used in any computation.

The constant term B_0 is found by computing the average amplitude of the actual wave over one cycle. Assuming any reference time $t = 0$ on the wave,

$$B_0 = \frac{1}{T}\int_0^T y(t)\,dt = \frac{1}{2\pi}\int_0^{2\pi} y(\omega t)\,d\omega t = \frac{1}{2\pi}\int_0^{2\pi} y(\theta)\,d\theta \quad (2)$$

where the angle, ωt, is replaced by θ. Since B_0 is a constant, or dc, term, it merely raises or lowers the entire wave and does not affect its shape.

The coefficients of the sine series are obtained by multiplying the wave $y(\theta)$ by $\sin n\theta$, integrating this product over a full cycle of the fundamental, and dividing the result by π. Thus,

$$A_n = \frac{1}{\pi}\int_0^{2\pi} y(\theta)\sin n\theta\,d\theta \quad (3)$$

The coefficients of the cosine terms are obtained in like manner except that $\cos n\theta$ replaces $\sin n\theta$. Thus,

$$B_n = \frac{1}{\pi}\int_0^{2\pi} y(\theta)\cos n\theta\,d\theta \quad (4)$$

If mathematical expressions describe $y(\theta)$, Eqs. (2), (3), and (4) give the coefficients of the series directly through analytical methods. If oscillograms or tabulated data describe $y(\theta)$, graphical or tabular forms of integration then are used

Effect of even harmonics. Figure 1 shows waves composed of a fundamental and a second harmonic only. In Fig. 1a, both ϕ_1 and ϕ_2 are zero. In Fig 1b, ϕ_1 is zero and ϕ_2 is $+\pi/2$ radians with respect to one cycle of the second harmonic. It is seen in the latter example that the negative part of the over-all wave is completely unlike the positive portion. Also, in general, these two sections will have different time intervals. Even harmonics give unsymmetrical waves.

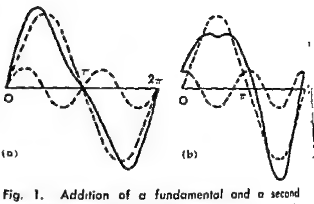


Fig. 1. Addition of a fundamental and a second harmonic. (a) $\phi_1 = 0$, $\phi_2 = 0$. (b) $\phi_1 = 0$, $\phi_2 = +\pi/2$ radians.

Effect of odd harmonics. Figure 2 shows waves composed of a fundamental and a third harmonic. In Fig. 2a, both ϕ_1 and ϕ_3 are zero. In Fig. 2b, ϕ_1 is zero and ϕ_3 is $+\pi$ radians with respect to one cycle of the third harmonic. In Fig 2c, ϕ_1 is zero and ϕ_3 is $+\pi/2$ radians on the third harmonic time scale. It is seen in these diagrams that the negative and positive parts of the over-all waves are alike and that both embrace π radians of the fundamental. Odd harmonics lead to symmetrical waves.

Symmetry. To determine symmetry of nonsine waves, the constant term B_0 is first removed. This means moving the wave down or up by the value of B_0. After this, if the wave from π to 2π is rotated about the horizontal axis and moved forward π radians, and then coincides exactly with the section from 0 to π, the total wave is said to have half-wave symmetry. Often the wave is said merely to be symmetrical. This means that in such cases $y(\theta + \pi) = -y(\theta)$. If, in turn, each half of the wave is symmetrical about the vertical axes of $\pi/2$, or $3\pi/2$, the wave is said to have quarter-wave symmetry as well. Half-wave and quarter-wave symmetry do not necessarily accompany each other.

All three waves of Fig. 2 have half-wave symmetry. The first two have quarter-wave symmetry also, but that of Fig. 2c does not. Waves having only a fundamental and odd harmonics will show half wave symmetry Conversely, half-wave symmetry indicates that only the fundamental, if it exists, and odd harmonics are present in the total wave. Half-wave symmetry permits Eqs. (3) and (4) to be integrated over the interval π, with the result multiplied by 2 Quarter-wave symmetry permits integration over one-quarter cycle of the fundamental, with the result multiplied by 4

Half-wave symmetry means that the fundamental and all odd harmonics may pass through their zero values at times quite distinct from each other. With quarter-wave symmetry the fundamental and the odd harmonics all pass through their zero values at the same time; therefore all phase angles ϕ_n are either zero or 180°

The wave of Fig. 1a has no symmetry of the kind discussed above. Waves containing only the fundamental and even harmonics, or even harmonics alone, are [illegible] .though half

agation in the medium in which the sound is moving. For example, a sound wave having a frequency of 1000 cycles per second would have a wavelength of approximately 1 ft in air, 4½ ft in water, and 17 ft in steel. The wavelength of electromagnetic waves depends on the velocity of light in the material in which the waves are traveling. *See* WAVE MOTION [W.J.G.]

Wavelength measurement

The wavelength of an oscillating electromagnetic wave depends upon the frequency of the oscillation and the velocity of propagation in the medium or the transmission system in which the wave is propagating (*see* WAVELENGTH; WAVEMETER). The wavelength in a transmission system is obtained by measuring the distance between successive wavefronts of equal phase. This measurement is most conveniently carried out in a standing-wave field in which interference occurs between forward-propagating and reverse-propagating, or reflected, waves. A distance equal to one-half wavelength exists between successive minima or maxima in the standing wave pattern.

The velocity of electric waves in free space is 299,792.5 kilometers per second or approximately 3×10^8 meters per second. The presence of any dielectric material (such as air) or any magnetic matter with permeability greater than unity will cause the wave to travel at a lower velocity. Also, the presence of various types of transmission lines and wave guides will affect the velocity. In general the group velocity, or velocity of propagation of the wavefront of a suddenly applied signal, is less than that in free space, but it is possible for the phase velocity to be greater than the value for free space. In particular, the phase velocity in a wave guide operated with waves near the cutoff frequency is greater than the phase velocity in free space.

Wavelength by frequency measurement. In order to specify the wavelength of an electromagnetic wave, it is essential to know the medium or device through which the wave is propagating. However, unless otherwise specified, it is general practice to quote the wavelength of an oscillatory electric wave as that in air, or free space. When this convention is used, all that is necessary is a frequency measurement in order to apply the relation

$$\lambda_0 = c/f$$

where c is the velocity of light, λ_0 is the free-space wavelength, and f is frequency.

In view of the direct relationship between frequency and wavelength the earliest method used for specifying the tuning point of a given radio wave was the wavelength of the wave. Early experimenters found that standing waves existed in space wherever reflections occurred. In such a measurement the distance between two maxima, or two minima, of a standing wave field is measured with a meter stick or measuring tape, the location of the field maxima being observed by moving a suitable detector unit to various points in the field.

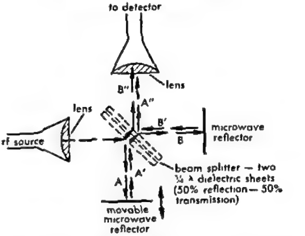


Fig. 1 Wavelength measurement by the Michelson interferometer used at millimeter wavelengths.

Interferometer methods. The standing-wave method of measurement is somewhat similar to the interferometer measurements used in optics. With shorter-wavelength radio waves it is possible to apply optically derived interferometer techniques directly to the measurement of wavelength.

An example of an interferometer used in the millimeter wavelength range is shown in Fig. 1. Essentially, a microwave beam is directed at a beam splitter, which splits the beam into two parts, A and B, by partial reflection. The A beam is reflected to a movable reflector and reflected again as A'. The beam splitter transmits part of this as A''. The transmitted part B of the original beam is reflected by a fixed microwave reflector as B'. This is partially reflected by the beam splitter as B''. The beams A'' and B'' combine to form standing waves, which are then detected. Movement of the movable reflector causes the position of the standing wave to move which causes the detected signal to pass through successive points of maximum and minimum amplitude. The distance between points of successive maxima or minima is one-half wavelength. This distance may be determined from the motion of the movable reflector.

Lecher wires. A more convenient, if less precise, measurement method for the determination of wavelength is the Lecher wire wavemeter (Fig. 2). With this simple device, wavelength is ...

dicated by absorption of signal being detected by an external detector, the variation in input power to oscillator or amplifier being checked, or by some similar method, such as the dc grid current of a negative-grid triode oscillator. The distance between two successive absorption maxima is one-half wavelength ($\lambda/2$); thus, by this length measurement, the wavelength is measured directly. *See* TRANSMISSION LINES; TRANSMISSION THEORY AND METHODS.

Tuned Circuits. At lower frequencies, inductance-capacitance resonant circuits may serve similar functions as absorption devices. With calibrated values of inductance L and capacitance C, it is

Example of nonsinusoidal waves. Assume a series circuit to have 8 ohms resistance and 15.91 millihenries inductance and that the impressed voltage is

$$e = 100 \sin 377t + 80 \sin 1131t$$

The problem is to calculate the rms voltage and current, the average power, and the power factor. The voltage is seen to have a fundamental component of 60 cycles ($f_1 = \omega_1/2\pi = 377/2\pi$) and a third harmonic of 180 cycles ($f_3 = \omega_3/2\pi = 1131/2\pi$).

At 60 cycles:

$X_{L_1} = 377 \times 0.01591$
$= 6.0$ ohms inductive reactance
$Z_1 = \sqrt{8^2 + 6^2} = 10$ ohms impedance
$I_1 = 100/10 = 10$ amp max fundamental current
$\theta_1 = \arctan (6/8) = 36.87°$

At 180 cycles:

$X_{L_3} = 3 \times 6 = 18$ ohms inductive reactance
$Z_3 = \sqrt{8^2 + 18^2} = 19.70$ ohms impedance
$I_3 = 80/19.7$
$= 4.06$ amp max third-harmonic current
$\theta_3 = \arctan\ (18/8)$
$= 66.06°$ ($= 22.02°$ on fundamental scale)

The equation for the current is

$$i = 10 \sin (377t - 36.87°) + 4.06 \sin (1131t - 22.02°)$$

$$E_{rms} = \sqrt{\frac{100^2}{2} + \frac{80^2}{2}} = 90.06 \text{ volts}$$

$$I_{rms} = \sqrt{\frac{10^2}{2} + \frac{4.06^2}{2}} = 7.63 \text{ amp}$$

Apparent power $= 90.06 \times 7.63 = 687$ volt-amperes
Average power $= I^2R = 7.63^2 \times 8 = 466$ watts

$$\text{Power factor} = \frac{466}{682} = 0.678$$

See ALTERNATING-CURRENT CIRCUIT THEORY.

[B.L.R., W.S.P.]

Bibliography: W. W. Lewis and C. F. Goodheart, *Basic Electric Circuit Theory*, 1958; M. B. Reed, *Alternating-current Circuit Theory*, 2d ed., 1956; H. H. Skilling, *Electrical Engineering Circuits*, 1957; K. Y. Tang, *Alternating Current Circuits*, 2d ed., 1951.

Waveform determination

The defining of a curve, or waveform, that represents the variation of the magnitude of a quantity with time. Waveform is determined either with oscillographs that display and record the waveform directly, or with wave analyzers that indicate the numerical values of amplitude, frequency, and sometimes phase angle of the harmonic components of a complex wave.

The measurement and control of waveform is of real concern to the electrical industry's engineers, because the transformers, motors, lighting circuits, ... equipment are designed to operate at maximum efficiency when the voltage waveform is a sine wave with a predetermined crest voltage. Departures from the specified waveform cause losses in efficiency. In radio communication systems the carrier waveform is sinusoidal, and deviations from the sine wave introduce noise that interferes with the intelligence being transmitted.

Waveforms are of two basic types: (1) periodic or continuous waves, and (2) aperiodic or transient waves. For examples and analysis of periodic waves *see* WAVEFORM; WAVEFORM, NONSINUSOIDAL. For aperiodic waves *see* TRANSIENT, ELECTRIC.

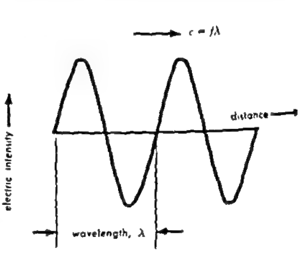
Two general classes of measurement devices are used to determine waveforms. The harmonic analyzer is widely used and is the most accurate for determining the waveform of a continuous wave. Any continuous wave can be defined by a Fourier series of sine waves, including a wave at fundamental frequency and waves at harmonics of this fundamental frequency. The harmonic analyzer indicates numerical values of the amplitude, frequency, and phase of the fundamental and each of the harmonics of the waveform under study. When required, the complete waveform may be constructed by graphically superimposing the component waves (*see* HARMONIC ANALYZER). The second type of measurement device, the oscillograph, is used to determine transient waveforms. The oscillograph may also be used to determine the waveform of continuous waves but with considerably less accuracy than that provided by the harmonic analyzer. *See* OSCILLOGRAPH; OSCILLOSCOPE, CATHODE-RAY; *see also* ELECTRICAL MEASUREMENTS.

[I.F.K.

Wavelength

The distance between two points on a wave which have the same value and the same rate of change of the value of a parameter, for example, electric intensity, characterizing the wave. The wavelength usually designated by the Greek letter λ, is equal to the speed of propagation c of the wave divided by the frequency of vibration f, that is, $\lambda = c/f$. *See* WAVE (PHYSICS).

The wavelength for a sound of a given frequency varies greatly, depending upon the speed of pro...



Wavelength λ and related

A modern resonant-cavity microwave wavemeter is shown in Fig 5. The dimensions of the cavity determine the resonant frequency of the wavemeter. A signal is fed in from either a coaxial line or wave guide, and energy is fed out to a suitable detector by a second coaxial line The cavity is tuned by means of a micrometer-driven plunger, which may be calibrated in terms of wavelength *See* COAXIAL CABLE; WAVE GUIDE [F.D L.]

Bibliography· E. L. Ginzton, *Microwave Measurements*, 1957; F. A Jenkins and H. E. White, *Fundamentals of Optics*, 3d ed , 1957 , C G. Montgomery, *Techniques of Microwave Measurements*, 1948, F. E. Terman and J. M. Pettit, *Electronic Measurements*, 2d ed , 1952

Wavelength standards

Accurately measured lengths of waves emitted by specified light sources for the purpose of obtaining the wavelengths in other spectra by interpolating between the standards

Spectra are produced by refraction in prisms or by diffraction from gratings (*see* SPECTROSCOPY) In either case, the spectra are usually photographed, and in order to identify or describe them reliably, it is necessary to measure the positions of the unknown spectral lines relative to known standards. It is customary to photograph the spectrum containing the standards either superposed or juxtaposed on the spectrum under investigation The wavelengths of hundreds of thousands of spectral lines have already been measured relative to a few hundred standards, and this will continue to be a major activity in spectroscopy.

Primary cadmium standard. In 1907, the International Union for Cooperation in Solar Research adopted the value 6438 4696 A as the primary standard for the wavelength of red radiation from cadmium measured relative to the meter. In 1910, the Union adopted 42 secondary standards in the arc spectrum of iron ranging from 4282.406 to 6494.985 A, the values being the means of three independent determinations In 1919, responsibility for standard wavelengths was assumed by the International Astronomical Union (IAU).

Secondary iron standards. In 1928, values ranging from 3370 786 to 6677.993 were adopted for 275 iron standards, and by 1937, the number had increased to 306 and the range to 2447 708 A. This international system of secondary standards of iron wavelengths, relative to the primary standard of cadmium, has been the foundation of spectroscopic measurement since 1928.

Between 1928 and 1937, individual values of iron wavelengths from 2100 794 to 10216 351 A had been reported, but since confirmatory observations were lacking the number of official standards could not be increased Additional measurements were made after 1950 so that by 1954 the means of three or more independent, concordant determinations existed for 575 iron wavelengths covering the spectrum from 2301 133 to 9372 904 A Instead of adopting these means as secondary standards of wavelength, a new procedure was followed by the International Astronomical Union—the best relative values of atomic energy levels were derived from these means, and then new secondary standards were calculated to eight figures from permitted combinations of the energy levels. If it is assumed that this combination principle of spectroscopy is exact, such a procedure reduces residual random errors of the original means and permits the calculation of additional wavelengths with equal precision Thus, in 1955, the IAU adopted calculated values of 1016 iron wavelengths ranging from 2084 1218 to 11973 067 A.

In actual practice, the positions of individual iron lines in a spectrum are rarely measured to seven figures because the lines are naturally wide and some are diffuse while others are partially self-reversed. The excessive line width is due mainly to the Doppler effect because of very high arc temperature Indeed, a temperature of 6300°K ± 10% has been deduced from the Doppler width of iron lines emitted by a 5-amp direct-current arc. Furthermore, collision broadening is present because the arc giving secondary standards of wavelength is operated at atmospheric pressure. (For further information on Doppler width and collision broadening of spectral lines, *see* ATOMIC STRUCTURE AND SPECTRA)

The only practical way of improving the iron standards is to replace the high temperature arc in air by sources operating at lower temperatures and pressures Two such sources have been developed, one is a hollow iron-cathode lamp containing neon at 3 mm of mercury pressure and the other is an electrodeless iron-halide lamp containing helium at 2 mm of mercury pressure and excited by microwaves The operating temperature of either lamp is near 800°K, and since the total gas and vapor pressures are only a few millimeters of mercury, the iron lines are less than one-third as wide as those from the international iron arc, and, consequently can be measured with greater accuracy

Provisional wavelengths of 283 iron lines ranging from 2457.5975 to 5709.3778 A, emitted by low-pressure and low-temperature sources, are collected in the 1958 Transactions of the IAU

Other secondary standards. Even this increase in accuracy of standard iron wavelengths will not satisfy future precision demands in spectroscopy. Iron is a light atom (mass 56) compared with cerium (mass 140) or thorium (mass 232) In comparable sources of radiation, the width of iron lines is double those of thorium lines Furthermore, iron has a simple spectrum compared with the spectra of rare-earth metals that emit two or three times as many lines In order to measure the wavelengths in very complex spectra with adequate accuracy, [illegible]

[illegible]

ley published the wavelengths of 222 thorium lines ranging from 3287.7885 to 6989 6562 A The combination principle indicates that the average error

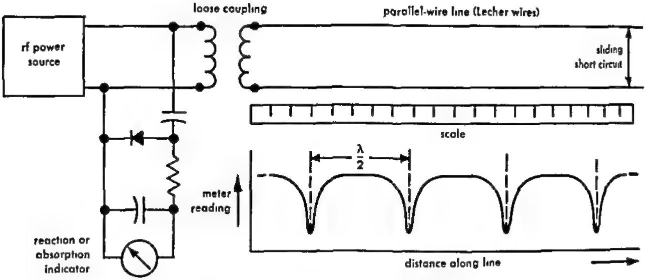


Fig. 2. Lecher-wire (resonant-line) wavemeter

possible to provide a scale calibrated in wavelength or frequency. For low-loss circuits the resonant frequency is $1/(2\pi\sqrt{LC})$. Figure 3 shows a schematic diagram of a simple absorption-type wavemeter. Wavemeters of this type, often constructed with the principal inductance L_1 as a plug-in coil, are used for frequency or wavelength measurement up to frequencies of approximately 1000 Mc.

A variation of the absorption wavemeter, known as a grid-dip meter or grid-dip oscillator, provides in one instrument an absorption wavemeter and a calibrated oscillator, which may be used to determine the resonant frequency (or wavelength) of a passive network, such as an antenna system or an *LC* tuned circuit, by observing the dip in grid current of the oscillator as the oscillator frequency is tuned to the network resonant frequency Figure 4 shows typical grid-dip oscillators covering frequencies from 2 2 to 1000 Mc *See* RESONANCE (ALTERNATING-CURRENT CIRCUITS).

Microwave wavemeters make use of resonant coaxial-line sections or cavities as tuned elements (*see* CAVITY RESONATOR). The two general types of microwave wavemeter are the absorption, or reaction, type and the transmission type. Wavemeters of low or medium selectivity are frequently used as coarse measuring devices to establish the general range of frequency of operation of a system before applying more refined and complex methods for accurate frequency checking

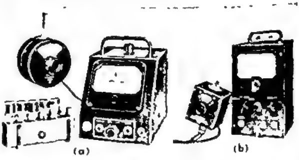


Fig 4. Grid-dip oscillator types of wavemeters (Note plug-in coils used with calibrated capacitors in each instrument) (a) For range from 2 2 to 400 Mc. (McGraw Edison Co) (b) For range from 300 to 1000 Mc. (Boonton Electronics Corp)

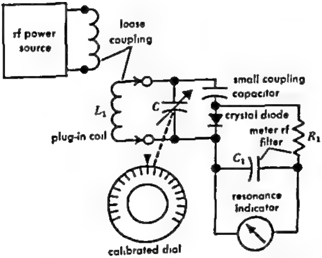


Fig. 3. Schematic diagram of inductance-capacitance type of absorption wavemeter (for frequencies between 50 kc and 1000 Mc, approximately).

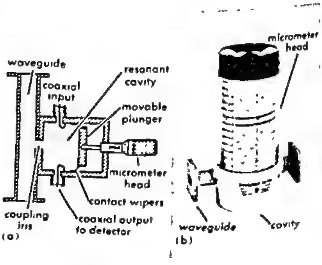


Fig 5 Resonant-cavity wavemeter. (a) Typical construction providing either coaxial or waveguide inputs (b) Commercial model. (Hewle ckard Co)

Other wave shapes of particular interest in electronics are the linearly increasing function of time, or ramp function (which if recurring at equally spaced periods is usually called the linear sawtooth waveform), the hyperbolic waveform, and the rectified sine wave (see RECTIFIER; SWEEP GENERATOR). Such recurrent waveforms are shaped by the combination of electronic switches or gating circuits, resistance-capacitance time-constants, and linear feedback amplifiers. See CLAMPING CIRCUIT; COINCIDENCE AMPLIFIER; ELECTRONIC WRITING; PULSE GENERATOR, SWITCH, ELECTRONIC; TIME-DELAY CIRCUITS, TRIGGER CIRCUIT. [G.M.C.]

Wax, animal and vegetable

Esters of high molecular-weight monohydroxy alcohols and high molecular-weight carboxylic acids. They are thus chemically different from fats and oils, from hydrocarbon or paraffin waxes, and from synthetic polyether waxes such as Carbowax, although all waxes possess the same characteristic physical properties of feel, consistency, melting point, and water insolubility. The materials commonly known as waxes are not chemically pure compounds, but are mixtures of substances of approximately similar physical properties, these mixtures may include hydrocarbon "waxes" together with ester waxes, and may be further contaminated with high-molecular-weight alcohols, acids, ketones, and ketonic alcohols. For this reason, the term *wax ester* is often used to distinguish the true ester type of wax. See ESTER; FAT AND OIL, NONEDIBLE.

The table presents the common, nonsynthetic waxes of commerce, together with their important physical characteristics.

Sources and structure. Waxes are found in the cuticles of plants, in cellular fabrications of bees (honeycombs) and other insects, as a coating on the leaves of many trees and on grasses, in certain portions of the bodies of various land and marine animals, in the envelopes of many seeds, on the hair of various animals, and associated with certain bacilli.

In many cases, the crude wax represents such a complex mixture of the above-mentioned substances that it has proved impossible to effect complete separation and identification of all the components. However, it can be stated that generally ester waxes are simple esters of alcohols (steroidal or open-chain), composed of even numbers of carbon atoms (from C_{24} to C_{36}) esterified with acids of similar carbon content. Occasionally, as in certain leaf waxes, only one alcohol is present; in such cases, it can be isolated in a state of purity. In general, then, the characterization of waxes for industrial or domestic use is physicochemical rather than structural and commonly makes use of the following terms:

The *solidification point* is the temperature at which the liquid form of the substance changes into the solid form. This temperature is not necessarily identical with the melting point.

The *acid value* (or *acid number*) is a measure of the quantity of free acids present in a body. In analyzing fats, oils, and waxes, it is defined as the

Common, nonsynthetic waxes*

Name	Specific gravity $^{15}/_{15}$°C	Solidification point, °C	Acid value	Saponification value	Iodine value
Animal wax					
Spermaceti (Kogia)	0.905–0.960	41–49	0.5–3.0	121–135	2.5–8.5
Spermaceti (*Physeter catodon*)	0.903–0.915	42–47	0.5–2.8	126–135	3.8–9.5
Wool fat	0.970–0.973	38–40	59.8	82–130	17–29
Insect wax					
Beeswax (Indian)	0.953–0.970	61–67	5.0–10.5	87–117	4–10.5
Beeswax (ordinary)	0.953–0.970	62–66	17.0–21.0	88–100	8–11
Chinese insect wax	0.950–0.970	80–85	1.9–8.9	78–93	1.0–2.5
Mineral wax					
Ceresine	0.900–0.920	56–82	0	0	4.0–8.0
Ozokerite	0.900–0.996	56–82	0	0	4.0–8.0
Semimineral wax					
Montan (domestic)	1.020–1.010	80–86	35.0–45.0	100–115	
Montan (Reibeck)	0.995–1.010	83–89	35.0–45.0	80–99	
Vegetable wax					
Candelilla	0.981–0.994	73–77	18.6–23.9	55–64	
Carnauba (crude)	0.994–1.010	86–90	3.0–8.5	75–89	
Carnauba (refined)	0.990–0.996	86–89	3.5–5.0	76–85	7.0–14.5
Carnauba (yellow)	0.990–0.996	86–88	1.5–2.5	75–86	
Esparto	0.985–0.995	75–79	22.0–27.0	58–72.5	7.0–15.0
Japan wax	0.970–0.998	49–56	4.0–15.0	210–235	4.0–15.0
Ouricury	0.990–1.010	86–89	12.0–18.8	88.0–95.8	

*Data based on N. A. Lange and G. M. Forker (eds.), *Handbook of Chemistry*, 9th ed., McGraw-Hill, 1956

in relative value is less than 1 part in 20,000,000. When more of these wavelengths are measured, especially for shorter waves, a system of secondary standards will be available to fill all present and foreseeable needs in the accurate description of highly dispersed and extremely complex spectra like those of cerium, praseodymium, uranium, or plutonium.

A few secondary standards of superior quality are available from neon, argon, or krypton excited in Geissler tubes. In 1955, the IAU adopted 30 neon wavelength standards (5400.5617–8082.4581 A), 56 argon standards (3947.5047–9784.5020 A), and 31 krypton standards (4273.9703–9784.5020 A). These standards do not extend into the ultraviolet and are too irregularly spaced to be useful in the measurement of complex spectra; some of them have been used in metrology to calibrate end gages.

Improvements in measurement. Further improvements in a primary standard of wavelength have come from the use of single even-mass isotopes in low-temperature sources, to eliminate effects due to the isotope shift and to reduce Doppler width (*see* ISOTOPE SHIFT). An electrodeless lamp containing mercury of mass 198 near 0°C is the most simple and convenient source; W. F. Meggers and K. G. Kessler have reported the wavelengths of 27 lines ranging from 2536.5064 to 5790 6626 A. Another source is a hot-cathode lamp containing krypton 86 immersed in nitrogen near 63°K. In 1957, the Advisory Committee on Redefining the Metre recommended that the meter be defined as exactly 1,650,763.73 wavelengths in vacuo of the radiation 6057 80211 A from krypton of mass 86.

Published data. Tables of international standard wavelengths appear in the IAU *Transactions*, vol. 9 and 10. Preliminary values of 222 standard thorium wavelengths have been published in the *Journal of Research* of the National Bureau of Standards, 61·95, 1958. Wavelengths of 27 lines from Meggers mercury-198 lamps have been presented in the *Journal* of the Optical Society of America, 40 737, 1950. [W F M.]

Wavellite

A hydrated phosphate of aluminum mineral with composition $Al_3(OH)_3(PO_4)_2 \cdot 5H_2O$, in which small amounts of fluorine and iron may substitute for hydroxyl group (OH) and aluminum (Al) respectively. Wavellite crystallizes in the orthorhombic system. The crystals are stout to long prismatic, but are rare. Wavellite commonly occurs as globular aggregates of fibrous structure and as encrusting and stalactitic masses. Crystals range in color from colorless and white to shades of blue, green, yellow, brown, and black.

Wavellite is a widespread secondary mineral occurring in small amounts in crevices of low-grade metamorphic aluminous rocks, in limonite, and in phosphate rock deposits. Found at many places in Europe and North America, it is also abundant in tin veins at Llallagua, Bolivia [W.R LO.]

Wavemeter

A device for measuring the geometrical spacing between successive surfaces of equal phase along an electromagnetic wave. To avoid instrument calibration problems due to the dependence of the phase velocity upon the particular transmission system under measurement, it is common procedure to calibrate wavemeters in terms of free-space wavelength, which is the ratio of the velocity of light (299,793,000 meters per second) divided by the frequency of the signal in cycles per second

At frequencies up to about 100 megacycles, wavemeters consist basically of a tuned *LC* circuit and a suitable resonance indicator not noticeably different from that originally used by H Hertz. The choice of a detector depends upon the power level of the signal and the desired accuracy. For power levels greater than several watts and for moderate accuracy, a miniature lamp bulb in series with the inductor *L* serves as a resonance indicator by glowing brightly when the induced current is a maximum. For low power levels or for higher accuracy the loading effect of the bulb cannot be tolerated Therefore a suitable vacuum-tube voltmeter is used to measure the capacitor voltage, which is a maximum at resonance. The capacitor is variable and calibrated in units of wavelength or frequency.

At higher frequencies it is necessary to employ well-defined transmission systems such as open-wire or coaxial lines and wave guides. Any open circuited or short-circuited section of transmission system can be adjusted in physical size to cause it to resonate at a given wavelength. From a construction standpoint short-circuited sections are preferred and may take the form of Lecher wires, coaxial wavemeters, and cavity resonators. Suitable electronic standing-wave detectors are employed to indicate when the wavemeter is tuned to resonance *See* CAVITY RESONATOR; STANDING-WAVE DETECTOR; WAVELENGTH MEASUREMENT. [R L.R]

Wave-shaping circuits

Electronic circuits used to create or modify a specified time-varying electrical quantity, usually voltage or current, using combinations of electronic devices, such as vacuum tubes or transistors, and circuit elements including resistors, capacitors, and inductors.

One such waveform is the square wave, in which a quantity such as voltage alternately assumes two discrete values during repeating periods of time (*see* SQUARE WAVE). Where each period is composed of two equal intervals, the wave can be obtained by amplifying a sinusoidal time-varying voltage and removing all but the section near the zero-voltage axis (*see* CLIPPING CIRCUIT; LIMITER CIRCUIT). Also, square waves of either equal or unequal intervals, sometimes referred to as rectangular waves, can be generated by multivibrators or various other forms of vacuum-tube or transistor switching or gating circuits. *See* MULTIVIBRATOR, RELAXATION OSCILLATOR.

Other wave shapes of particular interest in electronics are the linearly increasing function of time, or ramp function (which if recurring at equally spaced periods is usually called the linear sawtooth waveform), the hyperbolic waveform, and the rectified sine wave (see RECTIFIER; SWEEP GENERATOR). Such recurrent waveforms are shaped by the combination of electronic switches or gating circuits, resistance-capacitance time-constants, and linear feedback amplifiers. *See* CLAMPING CIRCUIT; COINCIDENCE AMPLIFIER; ELECTRONIC WRITING; PULSE GENERATOR; SWITCH, ELECTRONIC; TIME-DELAY CIRCUITS; TRIGGER CIRCUIT. [E.M.K.]

Wax, animal and vegetable

Esters of high-molecular-weight monohydroxy alcohols and high-molecular-weight carboxylic acids. They are thus chemically different from fats and oils, from hydrocarbon or paraffin waxes, and from synthetic polyether waxes such as Carbowax, although all waxes possess the same characteristic physical properties of feel, consistency, melting point, and water insolubility. The materials commonly known as waxes are not chemically pure compounds, but are mixtures of substances of approximately similar physical properties; these mixtures may include hydrocarbon "waxes" together with ester waxes, and may be further contaminated with high-molecular-weight alcohols, acids, ketones, and ketonic alcohols. For this reason, the term wax ester is often used to distinguish the true ester type of wax. *See* ESTER; FAT AND OIL, VEGETABLE.

The table presents the common, nonsynthetic waxes of commerce, together with their important physical characteristics.

Sources and structure. Waxes are found in the cuticles of plants, in cellular fabrications of bees (honeycombs) and other insects, as a coating on the leaves of many trees and on grasses, in certain portions of the bodies of various land and marine animals, in the envelopes of many seeds, on the hair of various animals, and associated with certain bacilli.

In many cases, the crude wax represents such a complex mixture of the above-mentioned substances that it has proved impossible to effect complete separation and identification of all the components. However, it can be stated that generally ester waxes are simple esters of alcohols (steroidal or open-chain), composed of even numbers of carbon atoms (from C_{24} to C_{34}) esterified with acids of similar carbon content. Occasionally, as in certain leaf waxes, only one alcohol is present; in such cases, it can be isolated in a state of purity. In general, then, the characterization of waxes for industrial or domestic use is physicochemical rather than structural and commonly makes use of the following terms:

The solidification point is the temperature at which the liquid form of the substance changes into the solid form. This temperature is not necessarily identical with the melting point.

The acid value (or acid number) is a measure *of the quantity of free acids present in a body. In* analyzing fats, oils, and waxes, it is defined as the

Common, nonsynthetic waxes*

Name	Specific gravity $^{15}/_{15}$°C	Solidification point, °C	Acid value	Saponification value	Iodine value
Animal wax					
Spermaceti (*Kogia*)	0.905–0.960	41–49	0.5–3.0	121–135	2.5–8.5
Spermaceti (*Physeter catodon*)	0.905–0.915	42–47	0.5–2.8	126–135	3.8–9.5
Wool fat	0.970–0.973	32–40	59.8	82–130	17–29
Insect wax					
Beeswax (Indian)	0.953–0.970	61–67	5.0–10.5	87–117	4–10.5
Beeswax (ordinary)	0.953–0.970	62–66	17.0–21.0	88–100	8–11
Chinese insect wax	0.950–0.970	80–85	1.9–8.9	78–93	1.0–2.5
Mineral wax					
Ceresine	0.900–0.920	56–82	0	0	4.0–8.0
Ozokerite	0.900–0.996	56–82	0	0	4.0–8.0
Semimineral wax					
Montan (domestic)	1.020–1.040	80–86	35.0–45.0	100–115	
Montan (Reibeck)	0.995–1.040	83–89	35.0–45.0	80–99	
Vegetable wax					
Candelilla	0.981–0.993	73–77	18.6–23.9	55–64	
Carnauba (crude)	0.994–1.019	86–90	3.0–8.5	75–89	
Carnauba (refined)	0.990–0.999	86–89	3.5–5.0	76–85	7.0–14.5
Carnauba (yellow)	0.990–0.995	86–88	1.5–2.5	75–85	
Esparto	0.975–0.995	75–79	22.0–27.0	58–72.5	7.0–15.0
Japan wax	0.970–0.998	49–56	4.0–15.0	210–225	4.0–15.0
Ouricury	0.990–1.010	86–89	12.0–13.8	83.0–95.8	

*Data based on N. A. Lange and G. M. Forker (eds.), *Handbook of Chemistry*, 9th ed., McGraw-Hill, 1956.

number of milligrams of potassium hydroxide required to neutralize the free fatty acids present in 1 g of substance

The saponification value (or number) is the number of milligrams of potassium hydroxide required for complete saponification of 1 g of the fat, oil, or wax.

The iodine value (or number) is a measure of the mean unsaturation present in the substance. It is defined as the number of grams of iodine absorbed by 100 g of sample. A high iodine value indicates a high degree of unsaturation of the material.

The Reichert-Meissel value (or number) is a measure of the amount of low-molecular-weight acids present in a sample. It is defined as the number of milliliters of decinormal alkali required for neutralization of the volatile, water-soluble fatty acids obtained from 5 5 g of fat, oil, or wax, by following a precise system of saponification, acidi*fication, and distillation. A large Reichert-Meissel* value indicates a large proportion of low-molecular-weight acids present in the sample.

Important waxes. Carnauba wax, a coating on the leaves of *Corypha cerifera*, a Brazilian palm, is one of the most important vegetable waxes. Chemically, it consists of a mixture of esters of type formula $CH_3(CH_2)_nCOO(CH_2)_{n+1}CH_3$, in which *n* is even and varies from 22 to 32, together with hydrocarbon waxes composed of odd-carbon molecules, C_{25} to C_{31} Carnauba wax is hard, has a relatively high solidification point (thus remaining solid in hot weather), is water-repellent, and can be polished to a fine luster. It finds extensive use in compounding floor and automobile polishes Japan wax, the fruit-coat of sumac berries, contains a high proportion of fat (glyceryl palmitate), together with about 6% of saturated, normal dibasic acids—mainly $C_{21}H_{42}(COOH)_2$ and $C_{20}H_{40}(COOH)_2$; the latter are supposed to be responsi*ble for the toughness of the wax and for its ability* to be kneaded without crumbling. Spermaceti wax, from the head of the sperm whale, consists mainly of cetyl myrist [illegible] cetyl laurate. [illegible] cetyl palmita [illegible] Beeswax, from [illegible] point intermediate between carnauba and spermaceti waxes; it consists mainly of esters formed from long, straight-chain acids and alcohols of 26 and 28 carbons. Wool wax *(degras or wool grease)*, obtained from the washing (scouring) of wool, is a highly complex mixture of wax esters, alcohols, and fatty acids Saponification of wool wax reveals the presence of a high proportion of branched-chain acids containing both even and odd numbers of carbons, together with α-hydroxy acids, normal and branched, so that *simple, straight-chain, high-*molecular-weight esters make up only a minor part of the wax. Wool wax forms stable emulsions with *up to 80%* water. Lanolin is purified wool wax or grease, used in salves and jellies and certain soaps Waxes of the envelopes of diphtheria, leprosy, and tuberculosis bacilli yield mycolic, phthioic, and tuberculostearic acids The last is D-(−)-10-methylstearic acid; phthioic acid is a mixture from which three acids have been identified, 3, 13, 19-trimethyltricosanoic acid (a C_{26} acid), mycolipenic acid (a C_{27} acid), and mycoceranic acid (a C_{31} acid). The mycolic acids are higher-molecular weight, branched-chain hydroxy acids. [E.B.K.]

Wax, petroleum

A substance produced primarily from the dewaxing of lubricating-oil fractions of petroleum. It may be of either the crystalline or microcrystalline type. The crystalline wax is produced from distillate lubricating fractions, whereas the microcrystalline wax is obtained from the residual lubricating fractions of the crude oil. The melting-point range for refined crystalline waxes is 120–150°F, while the petrolatum or microcrystalline waxes have melting points in the range of 150–175°F.

Petroleum waxes constitute approximately 90% of all of the wax used in industry today. The other 10% is comprised of vegetable and animal waxes which are used *primarily* as specialty waxes.

Petroleum wax and petrolatum production in the United States has varied depending upon the economic demand Current production is approximately 200,000,000 lb per year.

Uses. Petroleum wax has a wide variety of uses. For example, it is used to coat paper products, to *blend with other waxes for the manufacture of* candles, in the manufacture of electrical equipment and many polishes for home and industry, and as a source *material for oxidized* products which are being more widely used in industry The softer waxes, such as petroleum jelly, after proper purification, are being used as medicinal products.

The petrolata have many uses, principally as components of blended waxes used for coating paper, electrical insulation, and water- and rust proofing materials, as components of *crayons* and printing inks, and in the manufacture of specialties

In addition to melting point, other properties are important in establishing the quality and end use of waxes. These are tensile strength, hardness, stability, odor and taste, oil content, blocking and gloss characteristics, scuff resistance, and *sealing* strength

Purification of crude wax. The slack wax ob*tained from the primary dewaxing* operation on the lubricating fraction must be purified further in most cases before marketing The first operation in this purification is the removal of all or most of the oil associated with the wax The deoiling operation on the wax is conducted by either the sweating operation or the solvent deoiling operation

The old and widely used sweating process is, in reality, a fractional separation of crude wax (slack wax) obtained from a dewaxing process. The separation by melting points is accomplished in large pans or sweaters contained in a well insulated building which can be heated uniformly at a slow

rate to effect the melting-point separations. To accomplish this, molten slack wax is placed in the sweater pans and is cooled until a solid cake is formed. The sweater pans are then gradually warmed and the waxes are separated—the lowest-melting point fractions being separated as liquid fractions or cuts. As the temperature of the sweater is increased, the higher-melting-point waxes are removed. The cuts or fractions are then blended to provide the desired range of melting point. The melting points of the fractions vary from 110 to approximately 160°F. These fractions or blends are usually treated with sulfuric acid followed by clay filtration and deodorizing for the removal of impurities. The finished wax is molded into various shapes for commercial sale and distribution to industry. The commercial grades of paraffin wax are known as crude scale and fully refined waxes, and they vary in melting point range.

The solvent deoiling and fractionation process is a relatively new method of separating wax into its various melting-point fractions. It rapidly is replacing the sweating process because of its flexibility and wider range of application. This process is quite similar to the solvent dewaxing process. The slack wax to be deoiled or fractionated is dissolved in a suitable solvent and the mixture chilled to precipitate the desired melting-point grade of wax. The wax is separated by filtration from the solvent solution which retains the more soluble waxes and oil. Further chilling and filtration steps separate additional wax fractions. Additional treatment with acid and clay results in high quality waxes.

The petrolatum type of wax cannot be made by the sweating processes. These waxes are made directly from residual lubricating oil from which they are separated by either the solvent dewaxing process or by the centrifuge dewaxing method. Fractional separation of petrolatum into various melting points can be accomplished by the solvent deoiling process. One of the first methods of obtaining petrolatum was through the recovery of the wax bottoms that had settled out of wax-bearing crude oils during storage. See DEWAXING (PETROLEUM); PETROLEUM PROCESSING; PETROLEUM PRODUCTS.

[W.E.K.]

Waxwing

Any of three species of the Holarctic family Bombycillidae, all in the genus *Bombycilla*. Two similar species, *B. cedrorum*, the cedar waxwing, and *B. garrula*, the Bohemian waxwing, occur in the United States, where they are the only brown birds with a crest, or with a terminal yellow band on the tail. The larger Bohemian waxwing has white in the wing, and chestnut-red under tail coverts. The under tail coverts of the cedar waxwing are white, but there is no white on the wing. The cedar waxwing, by far the more common species, nests in the coniferous forests of southern Canada and the United States and winters from the central states southward into Mexico. Wandering flocks of these

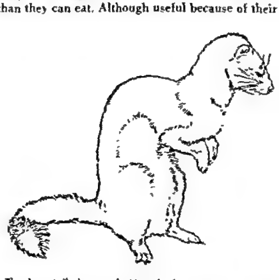
The cedar waxwing, *Bombycilla cedrorum*, length to 8 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

gentle, easily approached birds are characteristic of late winter and early spring. The Bohemian waxwing is more northerly in its distribution. Both species are erratic in their migrations. See PASSERIFORMES.

[J.D.B.]

Weasel

Any of a large number of carnivorous mammals of the genus *Mustela*, family Mustelidae, found throughout the Northern Hemisphere and in Africa and South America. Three species, including several distinct subspecies, are recognized in North America. Weasels are long, slender animals, with short, rounded ears and short legs, specially modified for pursuing burrowing animals into their dens. Weasels also prey upon any vertebrate of reasonable size, including squirrels, rabbits, and birds. They are among the few animals that will kill more than they can eat. Although useful because of their



The long-tailed weasel, *Mustela longicauda*; length to over 12 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

destruction of rodents, they can cause great damage in raids upon poultry houses.

Some weasels turn white in winter, this phase being known as ermine See CARNIVORA. [J.D.B.]

Weather

The state of the atmosphere, as determined by the simultaneous occurrence of several meteorological phenomena at a geographical locality or over broad areas of the earth. When such a collection of weather elements is part of an interrelated physical structure of the atmosphere, it is termed a weather system, and includes phenomena at all elevations above the ground. More popularly, weather refers to a certain state of the atmosphere as it affects man's activities on the earth's surface. In this sense, it is often taken to include such related phenomena as waves at sea and floods on land.

An orderly association of weather elements accompanying a typical weather system of the Northern Hemisphere may be illustrated by a large anticyclone, or high-pressure region. In such a "high," extending over an area of many thousands of square miles, the usually gentle winds circulate clockwise around the high-pressure center. This system often brings fair weather locally, which implies a bright sunny day with few clouds. The temperature may vary widely depending on season and time of day. On the other hand, a cyclone or low-pressure region is frequently associated with a dark cloudy sky with driving rain (or snow) and strong winds which circulate counterclockwise about a low-pressure center of the Northern Hemisphere.

A weather element is any individual physical feature of the atmosphere. At a given locality, at least seven such elements may be observed at any one time These are clouds, precipitation, temperature, humidity, wind, pressure, and visibility For discussion of a characteristic local combination of several elements, as they might be observed at a U.S Weather Bureau weather station, see WEATHER MAP. Each of these principal elements is divided into many subtypes.

The various forms of precipitation are included by international agreement among the hydrometeors, which comprise all the visible features in the atmosphere, besides clouds, that are due to water in its various forms. For convenience in processing weather data and information, this definition is made to include some phenomena not due to water, such as dust and smoke. Some of the more common hydrometeors include rain, snow, fog, hail, dew, and frost.

Both a physical (or genetic) and a descriptive classification of clouds and hydrometeors have been devised. The World Meteorological Organization, which among many other activities coordinates the taking of weather observations among the nations of the world, recognizes at least 36 cloud types and 100 classes of hydrometeors

Certain optical and electrical phenomena have long been observed among the weather elements These include lightning, aurora, and solar or lunar corona and halo. See AIR MASS; ATMOSPHERE; CLOUD; FRONT; PRECIPITATION (METEOROLOGY); STORM; WEATHER STATION; WIND [P.F.C.]

Bibliography: S. Petterssen, *Weather Analysis and Forecasting*, vol. 2, 2d ed., 1956; H. C. Willett and F. Sanders, *Descriptive Meteorology*, 2d ed. 1959.

Weather (forecasting and prediction)

Procedures for extrapolation of the future character of weather on the basis of present and past conditions. Accurate weather prediction requires a knowledge of the past state of the atmosphere, an understanding of the physical laws governing atmospheric behavior, and the availability of necessary technological aids for the rapid dissemination of meteorological information and the preparation of the forecast. The historical development of methods for forecasting the weather can be traced to innovations in these three areas. For introductory discussions of atmospheric science, see ATMOSPHERE; METEOROLOGY.

This article is divided into five sections. The first part emphasizes the current status of the whole field of weather forecasting and prediction, and concentrates upon short-range prediction, up to 48 hours in advance The second portion deals with long-range prediction. The third section considers statistical forecasting procedures. A fourth section summarizes the bases for the developing techniques of numerical prediction The last portion briefly characterizes the weather offices and centers through which are funneled the data of observation for analysis and processing into forecast and prediction For resumes of weather observation and data gathering, see METEOROLOGICAL INSTRUMENTATION; WEATHER STATION

DEVELOPMENT

Information on the state of the atmosphere has been greatly expanded since 1939, when the demands of aircraft in World War II led to the installation of radio-sounding stations and the development of weather reconnaissance systems. These sources have since been supplemented by radar aids, rockets, and, most recently, by satellites When these data are recorded on charts, the meteorologist has a three-dimensional picture of atmospheric structure. A series of such charts at 6-, 12-, or 24-hour intervals shows the development of weather systems in terms of the changes in wind, pressure, temperature, humidity, cloudiness, precipitation, and visibility. These maps permit an analysis of the dominant long-wave patterns (discussed under long-range forecast) with linear dimensions of 10^3–10^4 mi, the migratory cyclones and anticyclones with dimensions of 10^2–10^3 mi, and the weather conditions averaged over an area on the order of 10^4 mi.^{2} The resolution of weather detail is less exact over oceanic regions, in the polar areas, and in most tropical areas. Small scale weather systems such as land sea breezes, moun

tain valley winds, and convective showers cannot be depicted on the conventional weather map and can be studied only by means of a dense network of observing stations Radar information gives precipitation detail within the 10^4-mi^2 area

Data to forecasting. The step from data to forecast is achieved by a variety of methods which may be classified as follows.

Semiempirical techniques. The forecaster first synthesizes the raw information into dynamically meaningful models of the atmosphere (as best exemplified by the polar-front model of J. Bjerknes and H Solberg). Then, by a combination of methods including the extrapolation of past trends, expected changes based upon qualitative physical reasoning, and recollection of the behavior of similar situations in the past, he arrives at an estimate of the position of the prominent features on tomorrow's weather charts These features include the location and intensity of cyclones, anticyclones, fronts, upper level pressure ridges and

... questionable results The details of weather

... vection, radiation, topographical influences, and expected stability changes

The success of these techniques is limited by their semiempirical character and by the ability of an individual to handle the tremendous mass of significant information Data are now so voluminous that this method of prediction utilizes only a fraction of the available information, and it becomes somewhat a matter of personal choice as to the selection process

Numerical methods. The advent of an expanded network of weather observations and of high-speed computers has greatly stimulated meteorological research so that forecasting is passing from the pre-1945 qualitative phase to a quantitative era in which predictions are based upon computation, guided by physical principles These methods are discussed in the section on the numerical prediction of weather and the statistical prediction of weather

At present the large-scale aspects of the atmospheric flow pattern, especially in the middle troposphere, are most accurately predicted by numerical methods The detailed features of the weather are deduced by the methods discussed in the preceding section or by statistical procedures

Evaluating forecasting. To analyze the accuracy of weather forecasts, it is necessary to recognize the statistical nature of the element being predicted Precipitation, even when it covers wide areas, is rarely uniform in intensity Small convective showers, which develop and dissipate rapidly, are often imbedded in the general rain or snow and cause significant variations over distances of the order of 10 mi and over time periods of less than 1 hour. The shower or thunderstorm cells which are associated with air-mass showers or with cold fronts exhibit a maximum variability in time and space. This turbulent behavior presents a major forecasting problem. Consequently, the prediction must give wide range to the estimates of precipitation intensity in order to cover the probable variability over an area on the order of 10^4 mi.^{2} Only in those areas where the local variations can be attributed to orographical influences is it possible to present a more definitive estimate.

In contrast to the variability of precipitation, the turbulent fluctuations of wind and temperature are so rapid that they are not of general interest to the forecaster or the public. The specialized problem of predicting atmospheric pollution is an exception Local variations in temperature and wind within the 1-hour, 10-mi scale can be attributed to such well-understood influences as the nature of the underlying surface, proximity to a water area, and a valley or mountain effect. Hence, detailed forecasts of these elements can be made with considerable reliability

The final consideration with all types of forecast is the length of the time step, the larger the scale of the atmospheric system, the more persistent the phenomenon. For example, an individual summer shower has a life span of approximately 1 hour; a cyclone is ordinarily identifiable for at least 3 days, and a particular long-wave pattern may persist for weeks At present, the major problem in weather prediction is the forecasting of a new development; it may be as difficult to pinpoint a summer shower a few hours in advance as to predict, a week in advance, the broad-scale features of the weather associated with a long-wave pattern For similar reasons, the accuracy with which the weather may be predicted in detail decreases rapidly with the time elapsed since the observations were made.

Probable future developments. This article is being written during a period of rapid development in the application of physical principles to the forecast problem and in the evolution of numerical and statistical techniques Some immediate improvement in short-range forecasting can be expected with the increase in speed of transmission of weather observations and the reduction in the length of time required to process the data, due to automation. Real improvement must await new numerical or statistical methods compatible with the developing technology in the computer area.

[J M.AU.]

LONG-RANGE FORECASTING

Long-range weather forecasts are of two types. Medium or extended range forecasts cover periods of from 48 hours to a week in advance. Forecasts for longer periods generally extend over periods of a month, a season, or possibly two or more seasons in advance. Although meteorologists have been working on long-range prediction problems for over 100 years, the current degree of accuracy is small for all predictions exceeding a week in ad-

vance. This seems particularly true when examined against rigid statistical controls, such as climatological probability. There is little or no evidence to indicate any sustained success for forecasts embracing periods of more than a month in advance. The reason for such limited ability lies in the utter complexity of the atmosphere's behavior—the vicissitudes of a compressible fluid responding to changing external stimuli, such as the sun, and changing characteristics of the earth's surface, both in space and time.

Medium or extended ranges. Scientific methods of extended-range prediction take for granted that the further out in time the forecast is projected, the more general must be the nature of the prediction. Short-range forecasts for 48 hours or less in advance specify the detail of weather in space and in time. Medium-range forecasts for a week in advance cover average conditions and trends within the week in intervals of a couple of days. Forecasts for a month or more, however, can indicate only the broad-scale (for example, areas of several hundred thousand square miles) features of average or prevailing weather. The latter are usually expressed in terms of departures from seasonal norms for elements such as temperature and precipitation

Medium-range forecast methods are apt to use one or a combination of dynamic, statistical, and synoptic techniques. Dynamic methods capitalize upon the best physical knowledge of meteorological phenomena. Statistical methods employ empirically derived equations as substitutes for physical knowledge. In the synoptic technique, various hemispheric wind and weather charts are surveyed and interpreted by an experienced meteorologist. The core of modern dynamic methods lies in the principle that the vertical component of absolute vorticity remains fairly constant as air columns of the middle troposphere move from one area to another (discussed under numerical prediction of weather). When instantaneous wind and pressure charts for midtropospheric levels are averaged in time or space, certain small-scale perturbations, including short waves or vortices in the horizontal, are suppressed. What remains are smooth, long, or planetary waves which in effect comprise a special class of motions; they are not only of larger scale (often being composed of a family of cyclones or anticyclones), but they also evolve more slowly than the individual wind charts from which they are constructed. The planetary waves which these time-averaged charts reveal are responsible for variations in the position and intensity of the well-known sea-level centers of action (like the Bermuda High, one of the subtropical oceanic highs) which largely determine prevailing weather abnormalities. For purposes of extended forecasting, the averaging process is performed on past (observed) data as well as on numerically predicted charts. Various methods of comparison of such time-averaged charts enable the synoptician to assay the continuity and trend of large-scale systems and to extrapolate them into the future for some reasonable period. It turns out that dynamic methods of prediction may also be used with some success on time-averaged charts, but the physical reasons for this are not clear.

Dynamic methods employing the vorticity concept may be improved in practice by directly inserting corrections stemming from empirical studies Thus statistical factors, which approximate the net effect of mountains and the heating of air currents in transit from land to ocean, are derived from vast accumulation of past data and inserted objectively. These are at best probabilistic corrections

Procedures for extended forecasting vary around the world largely in accordance with facilities and availability of scientific manpower. Many countries do not have available the high-speed computing equipment necessary to prepare the dynamic component of the forecast, nor have they even the statistical components. In these cases extended forecasts are either not prepared at all or are made by educated synoptic guesswork.

After predictions of average planetary wind flows at upper levels have been made, it is possible to infer the accompanying types of weather in different areas as well as to estimate the general regions for breeding and movement of storms and air masses. Here again the statistics of the motions and weather are much more predictable than the day-to-day detail. In fact, the translation of average wind circulation into average weather is amenable to statistical stratification procedures and is fairly, though not completely, objective.

Longer-period forecasting. For periods more than a week in advance, methods of forecasting rely more upon statistics and less upon physical reasoning. Concentrated efforts to explore the physics of long-range weather phenomena began about 1955 as a result of increased availability of computing facilities and hemisphere-wide data coverage, particularly from the upper air But even now, attempts are being made to prepare forecasts a month ahead by employing dynamic principles in conjunction with statistical and synoptic techniques. For these purposes, another class of mean motions is defined by construction of mean maps for 30-day periods Although real understanding of these methods is remote, recent experiments indicate that such objective, machine-produced prognoses are helpful and may in time supplant more subjective techniques.

Another less expensive and less time-consuming method of longer-range prediction involves the use of statistical analogs. The historical files of weather maps for past periods (mean maps also if desired) are searched and a wind and weather pattern is sought which is as similar as possible to the one which has been operative, with the assumption that what transpired in the earlier case will repeat. For best results the analogy should be good for large areas of the hemisphere, should hold for upper levels as well as for sea level, and should

tand up for a sequence of periods preceding the forecast The logic of this method is appealing: similar patterns under the same stimuli (such as season of the year) repeat themselves. However, the relatively short span of time for which meteorological records have been kept makes it difficult to find good analogs.

For periods beyond a month, statistical techniques seem to be the only ones sufficiently accurate for use at present; even here there is some question as to whether the samples of data (length of long period record) are adequate to assure the stability of discovered relationships. Thus many claims of long period correlations between meteorological events in remote regions and for long time lags may fail miserably when applied to independent data Similarly, attempts to use indices of solar activity as guides in long range weather prediction [illegible]

[illegible] is inadequate for purposes of prediction. [J.N.]

STATISTICAL WEATHER FORECASTING

Statistical weather forecasting is the prediction of weather by rules *based upon the statistics of weather behavior* *A prediction* may state the expected value of a specific weather element, such as a wind speed, or the probability of occurrence of a specific weather event, *such as a thunderstorm. In* the former case *the prediction* is understood to contain an error whose probable value may or may not be stated. The choice of the form of prediction may depend upon the intended *audience*; for example, the statement *that* there are two chances in ten that tonight's temperature will fall below 32°F might aid a *fruit* grower more than the statement that tonight's expected minimum temperature will be 36°F.

Basic premises. Statistical forecasting is based upon the premise that *the future* world-wide state of the atmosphere *is* determined, at least approximately, by the present state, *together* with the intervening influences of the sun and the underlying ocean and *land*, according to immutable physical laws. In *theory*, forecasting is equivalent to solving the *equations representing these laws*, but the equations are rather *intractable, and because there* are vast *gaps* between observing stations, the present weather *is only* partially known It is sometimes more *feasible* to ascertain how future weather must evolve from present weather by studying how the observed portions of the atmosphere have previously behaved.

Prediction rules established from such study often relate future weather to *present* and past weather, *instead* of present weather alone, since past knowledge may partially compensate for incomplete present knowledge *A rule* is commonly expressed as a mathematical *formula* Sometimes the same information is more conveniently presented as a graph or a table

A rule established for one location does not generally apply at another location. For example, a table established for San Francisco, showing the probability of occurrence of nighttime fog following various combinations of midafternoon temperature and relative humidity, would not be valid for predicting fog in New York. A new rule is usually needed for each new weather prediction in any particular area.

Statistical procedures. A general kind of procedure is used to establish a formula The meteorologist chooses a set of weather elements for what may be termed predictors and selects, commonly from past records, a set of data consisting of corresponding observed values of the predictors and the predictand. As the next step, he chooses a mathematical form with a limited number of degrees of freedom, ordinarily appearing as undetermined constants, and restricts the formula to this form He specifies a process, ordinarily the minimization of the sum of squares of the prediction errors, by which the chosen data shall determine the constants Evaluating the constants is then an objective and usually routine mathematical task.

The meteorologist may modify this procedure and classify combinations of values of the chosen predictors into categories He may then construct a table by choosing, for each category, the average observed value of the predictand as the expected value or, alternatively, by choosing the observed frequency of occurrence as the probability.

The preparation of a forecast once the rule is established is objective and usually simple. The forecaster evaluates the formula after introducing the appropriate numerical values of the predictors, or reads the forecast from the appropriate location in the graph or table.

The data selected for establishing a formula comprise a finite sample of the total history of the weather The formula is likely to succeed, when applied to future weather, only if the number of degrees of freedom is small compared to the number of values of each predictor *in the sample* since virtually any finite set of numbers will fit a sufficiently complicated formula

Ideally the sample should be made very large. When this is not feasible because of the excessive labor involved or the absence of extensive past records, the degrees of freedom must be restricted This is accomplished by limiting the number of predictors or restricting the formula to a *more* highly specialized form Meteorological experience or physical reasoning should be used as a guide since a blind choice of *predictors*, or of a mathematical form, is unlikely to yield a *successful* formula

Statistical linear regression. The simplest mathematical form, and the one whose theory is *most* highly developed is the linear formula. When *many* predictors have been chosen, the number *may* be reduced either by factor analysis, which selects a few linear combinations of the predictors in *order of their* ability to represent all the predictors, or

by a procedure which selects a few predictors in order of their independent contribution to the prediction. Widespread investigation of linear formulas has followed the advent of high-speed electronic computing machines. *See* STATISTICS.

Appropriate nonlinear formulas are theoretically superior to linear formulas, but since they usually involve many degrees of freedom, they are more difficult to discover.

Statistical methods are highly suitable for predicting special local phenomena such as the occurrence of fog. For preparing prognostic weather maps one or two days in advance, statistical formulas show positive usefulness, but are frequently inferior to conventional subjective forecasts. For forecasting several days in advance, linear statistical formulas show a slight positive utility and compare favorably with other methods. [E.N.L.]

NUMERICAL WEATHER PREDICTION

Numerical weather prediction is the prediction of weather phenomena by the numerical solution of the equations governing the motion and changes of condition of the atmosphere. More generally, the term applies to any numerical solution or analysis of the atmospheric equations of motion.

The laws of motion of the atmosphere may be expressed as a set of partial differential equations relating the instantaneous rates of change of the meteorological variables to their instantaneous distribution in space. These are developed in dynamical meteorology (*see* METEOROLOGY). A prediction for a finite time interval is obtained by summing the succession of infinitesimal time changes of the meteorological variables, each of which is determined by their distribution at the preceding instant of time. While this process of integration may be carried out in principle, the nonlinearity of the equations and the complexity and multiplicity of the data make it impossible in practice. Instead one must resort to finite-difference approximation techniques in which successive changes in the variables are calculated for small, but finite, time intervals at a finite grid of points spanning part or all of the atmosphere. Even so, the amount of computation is vast, and numerical weather prediction remained only a dream until the advent of the modern high-speed electronic computing machine. These machines are capable of performing the millions of arithmetic operations involved with a minimum of human labor and in an economically feasible time span. At the present time numerical methods are gradually replacing the earlier, more subjective methods of weather prediction in many government weather services. This is particularly true in the preparation of prognoses for large areas. The detailed prediction of local weather phenomena has not yet benefited greatly from the use of numerico-dynamical methods, as indicated in the general section on weather forecasting.

Short-range numerical prediction. By its nature, the accuracy of numerical weather prediction depends on (1) an understanding of the laws of atmospheric behavior, (2) the ability to measure the instantaneous state of the atmosphere, and (3) the accuracy with which the solutions of the continuous equations of motion are approximated by finite-difference means. So far, the greatest success has been achieved in predicting the motion of the large-scale (>1000 mi) pressure systems in the atmosphere for relatively short periods of time (1–3 days). For such space and time scales, the as yet poorly understood energy sources and frictional dissipative forces may be largely ignored, and one may use rather coarse space grids.

The large-scale motions are characterized by their properties of being quasi-static, quasi-geostrophic, and horizontally quasi-nondivergent, as discussed in another article (*see* METEOROLOGY). These properties may be used to simplify the equations of motion by filtering out the motions which have little meteorological importance, such as sound and gravity waves. The resulting equations then become, in some cases, more amenable to numerical treatment.

A simple illustration of the methods employed for numerical weather prediction is given by the following example. Consider a homogeneous, incompressible, frictionless fluid moving over a rotating, gravitating plane in such a manner that the horizontal velocity does not vary with height. For quasi-static flow the equations of motion are

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -g\frac{\partial h}{\partial x} + 2\omega v$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -g\frac{\partial h}{\partial y} - 2\omega u$$

and the equation of mass conservation is

$$\frac{\partial h}{\partial t} + u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y} = -h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)$$

where u and v are the velocity components in the directions of the horizontal rectangular coordinates x and y, t is the time, g is the acceleration of gravity, ω is the angular speed of rotation, and h is the height of the free surface of the fluid. Let the variables u, v, and h be defined at the points $x = i\,\Delta x$, $y = j\,\Delta x$ ($i = 0, 1, 2, \ldots, I$; $j = 0, 1, 2, \ldots, J$) and at the times $t = k\,\Delta t$ ($k = 0, 1, 2, \ldots, K$) and denote quantities at these points and times by the subscripts i, j, and k. Derivatives such as $\partial u/\partial t$ and $\partial u/\partial x$ may be approximated by the central difference quotients

$$\frac{\Delta_t u_{i,j}}{2\,\Delta t} = \frac{u_{i,j,k+1} - u_{i,j,k-1}}{2\,\Delta t}$$

and

$$\frac{\Delta_i u_{j,k}}{2\,\Delta x} = \frac{u_{i+1,j,k} - u_{i-1,j,k}}{2\,\Delta x}$$

respectively. In this way one obtains the following finite-difference analogs of the continuous equations:

$$u_{i,j,k+1} = u_{i,j,k-1} - \frac{\Delta t}{\Delta x}\left(u_{i,j,k}\,\Delta_i u_{j,k} + v_{i,j,k}\,\Delta_j u_{i,k} + g\,\Delta_i h_{j,k}\right) + 4\omega v_{i,j,k}\,\Delta t$$

$$v_{i,j,k+1} = v_{i,j,k-1} - \frac{\Delta t}{\Delta x}\left(u_{i,j,k}\,\Delta_i v_{j,k} + v_{i,j,k}\,\Delta_j v_{i,k} + g\,\Delta_j h_{i,k}\right) - 4\omega u_{i,j,k}\,\Delta t$$

$$h_{i,j,k+1} = h_{i,j,k-1} - \frac{\Delta t}{\Delta x}\left[u_{i,j,k}\,\Delta_i h_{j,k} + v_{i,j,k}\,\Delta_j h_{i,k} + h_{i,j,k}(\Delta_i u_{j,k} + \Delta_j v_{i,k})\right]$$

which give u, v, and h at the time $(k+1)\,\Delta t$ in terms of u, v, and h at the times $k\,\Delta t$ and $(k-1)\,\Delta t$. It is then possible to calculate u, v, and h at any time by iterative application of the above equations.

It may be shown, however, that the solution of the finite-difference equations will not converge to the solution of the continuous equations unless the criterion, $\Delta s/\Delta t > c\sqrt{2}$ is satisfied, where c is the maximum value of the speed of long gravity waves, $\sqrt{gh}$. Under circumstances comparable to those in the atmosphere one finds that Δt must be so small that a 24-hour prediction requires some 200 time steps and approximately 10,000,000 multiplications for an area the size of the earth's surface. The computing time on a machine with a multiplication speed of 100 μsec, an addition speed of 10 μsec, and a memory access time of 10 μsec would be about 30 min. The magnitude of the computational task may be comprehended from the fact that the more accurate atmospheric models now envisaged will require some 100–1000 times this amount of computation.

A saving of time is accomplished by utilizing the quasi-nondivergent property of the large-scale atmospheric motions. If, in the above example, the horizontal divergence $\partial u/\partial x + \partial v/\partial y$ is set equal to zero one finds that the motion is completely described by the equation for the conservation of the vertical component of absolute vorticity, as developed in another article (*see* METEOROLOGY). The solution of this equation may be obtained in far fewer time steps since gravity wave motions are filtered out by this constraint and the velocity c in the Courant-Friedrichs-Lewy criterion becomes merely the maximum particle velocity instead of the much greater gravity wave speed.

Cloud and precipitation prediction. If, to the standard dynamical variables u, v, w, p, and ρ, one adds a sixth variable, the density of water vapor, it becomes possible to predict clouds and precipitation as well as the air motion. When a parcel of air containing a fixed quantity of water vapor ascends, it expands adiabatically and cools until it becomes saturated. Continued ascent produces clouds and precipitation.

To incorporate these effects into a numerical prediction schema one adds the equation governing the rate of change of specific humidity r

$$\frac{Dr}{Dt} = \frac{\partial r}{\partial t} + u\frac{\partial r}{\partial x} + v\frac{\partial r}{\partial y} + w\frac{\partial r}{\partial z} = S$$

where S represents a source or sink of moisture. Then it is necessary also to include as a heat source in the thermodynamic energy equation a term which represents the time rate of release of the latent heat of condensation of water vapor. The most successful predictions made by this method are obtained in regions of strong rising motion, whether induced by forced orographic ascent or by horizontal convergence in well-developed depressions. The involved physics and mechanics of the convective cloud-formation process make more difficult the prediction of convective cloud and showery precipitation.

Extended-range numerical prediction. The extension of numerical predictions to long time intervals requires a more accurate knowledge than now exists of the energy transfer and turbulent dissipative processes within the atmosphere and at the air-earth boundary, as well as greatly augmented computing-machine speeds and capacities. However, predictions of mean conditions over large areas may well become possible before such developments have taken place, for it is now possible to incorporate into the prediction equations estimates of the energy sources and sinks—estimates which may be inaccurate in detail but correct in the mean. Several mathematical experiments involving such simplified energy sources have yielded predictions of mean circulations that strongly resemble those of the atmosphere.

Numerical calculation of climate. The above-mentioned experiments lead to a hope that it will be possible to explain the principal features of the earth's climate, that is, the average state of the weather, well before it becomes possible to predict the daily fluctuations of weather for extended periods. Should these hopes be realized it would then become possible to undertake a rational analysis of paleoclimatic variation and changes induced by artificial means. If the existing climate could be understood from a knowledge of the existing energy sources, atmospheric constituents, and earth surface characteristics, it might also be possible to predict the effects on the climate of natural or artificial modifications in one or more of these elements. [J.G.C.]

CENTERS AND OFFICES OF FORECASTING

Weather forecasts for all parts of the United States are prepared by the Weather Bureau of the U.S. Department of Commerce. Various phases of the forecast work are performed at three working levels: (1) the National Meteorological Center, (2) forecast centers, and (3) local offices.

National meteorological center. The National Meteorological Center, located in Suitland, Md., near Washington, D.C., collects weather observa-

Fig. 1. Weather data is received on teletypewriters recording on tape that is then fed into a component for converting the information into punch-card records for further processing (U.S. Weather Bureau)

tions, prepares weather charts, and issues forecasts on the future state of the atmosphere on a hemispheric scale. Weather observations taken throughout the Northern Hemisphere are sent to the Center at frequent intervals by landline and radioteletypewriter circuits and other rapid communication systems (see Figs. 1 and 4). Reports of weather elements measured near the surface of the earth throughout the Northern Hemisphere are collected four times a day. Every 24 hours the Center receives 25,500 surface reports of which 22,300 are from land stations and about 3,200 from ships at sea. Reports of upper air conditions are received daily from 900 pilot balloon observations of wind direction and speed; 980 radiosonde observations of upper air pressure, temperature, and humidity; and 500 rawinsonde (radar wind radiosonde) observations of pressure, temperature, humidity, and winds. The Center also receives some 900 reports from commercial aircraft in flight and 400 reports from scheduled military weather reconnaissance.

These reports are plotted on hemispheric charts for altitudes ranging from sea level to about 60,000 ft and are then synoptically analyzed. After analyzing the broad-scale weather patterns, the National Meteorological Center prepares forecast charts showing the changes in the large- and medium-scale features that are to be expected over the next 1–2 days. Both the analyzed charts and the forecast charts are transmitted by facsimile to forecast centers and local offices of the Weather Bureau, to other governmental agencies, and to private meteorological concerns. Certain of the charts are sent by radio facsimile to ships at sea and to foreign countries.

Most of these charts are still produced by hand, but several are now based on the output of a high-speed electronic computer. In some cases the numerical, machine-made chart is transmitted without change.

In addition to the daily forecast charts, the National Meteorological Center prepares 5-day and 30-day forecasts and outlooks. At present the 5-day forecasts take the form of forecast charts of air flow in the Northern Hemisphere and frontal patterns at sea level (one for each of the five days); they also include charts of the expected departures from normal of precipitation and temperature. The 30-day outlooks are issued as charts showing the expected mean monthly air flow over the Northern Hemisphere and the expected 30-day departures from normal of precipitation and temperature.

Forecast centers. The forecast centers make use of the facsimile charts received from the National Meteorological Center in their preparation of forecasts. The area served by a forecast center averages about 120,000 mi² or two average-sized states. These centers also receive additional detailed data in the form of hourly observations taken from the North American continent and 6-hourly reports from ships at sea.

Area refinements and warnings. Information received on the facsimile charts is supplemented by plotting and analysis of such additional data as may be required to introduce refinements in the shorter-range periods of the forecasts for the respective districts. Radar weather information is also utilized. A 24-hour vigil of all meteorological phenomena which might endanger life and property within a region is maintained; high winds, thunderstorms, blizzards, heavy snows, cold waves, killing frosts, fog, and heavy rain are among the weather phenomena for which warnings are issued by the forecast center to local offices.

Forecasts. In addition to issuing warnings, the forecast center makes regular forecasts at 6-hour intervals for issue to radio and the press. These forecasts contain information regarding expected cloudiness, precipitation, temperature, wind, humidity, and other weather factors of interest to the public. The center prepares 5-day forecasts on Mondays, Wednesdays, and Fridays for relatively

Fig. 2. Weather Bureau meteorologist at computer console selecting punch cards which contain meteorological data to be processed according to a program for analyzing weather patterns. (Joint Numerical Weather Prediction Unit, U.S. Bureau)

Fig 3 Meteorologist at Washington National Airport studying latest analyzed weather chart being transmitted by the National Meteorological Center via a facsimile network to several hundred stations. (U.S. Weather Bureau)

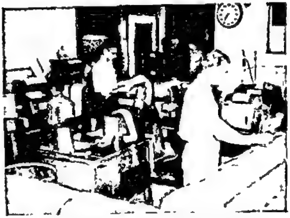


Fig 4 Communications room of Weather Bureau office at Washington National Airport where thousands of weather reports and forecasts are received daily via weather teletypewriter circuits connecting all parts of the world (U.S. Weather Bureau)

broad areas using guidance charts received from the National Meteorological Center.

Aviation and other special services The basic weather service to aviation consists of warnings and forecasts prepared at forecast centers. A special unit of aviation meteorologists issues several different types of forecasts at regular intervals These include aviation terminal forecasts, area forecasts, upper wind forecasts, and advisories of conditions considered hazardous to aircraft in flight. Through teletypewriter communications, these aviation weather forecasts are transmitted to local Weather Bureau offices, FAA traffic control and communication stations, Air Force, Navy, Army, and other government agencies operating aircraft, and to many private businesses directly concerned with aviation.

Certain of the forecast centers are assigned special functions such as issuing hurricane warnings, severe local storm warnings, international aviation forecasting, ocean area forecasts for merchant shipping, and river and flood forecasting The hurricane warning centers issue hurricane warnings and advisories giving the storm's position, intensity, direction and rate of movement, and a statement of the effects to be experienced, such as the onset of high tides or destructive winds with their direction and probable duration The Severe Local Storm Forecast Center (SELS) at Kansas City, Mo, specializes in forecasting areas (about 100 by 200 mi) likely to experience tornadoes, hailstorms, or severe thunderstorms. The international aviation centers forecast weather and upper air conditions for international flights originating in the United States The marine centers prepare forecasts of surface weather and sea conditions over the high seas Separate and distinct from the weather forecasting centers, the Weather Bureau also maintains river forecast centers A river forecast center is responsible for maintaining a continuous river watch and issuing forecasts of river and flood conditions. Advance warnings minimize the damaging effects of flooding by providing time for the evacuation of people and movable property.

Local offices. Local Weather Bureau offices throughout the nation provide reports, advisories, warnings, forecasts, and information services to the public in their respective communities and surrounding areas Most of this information is adapted from forecasts and warnings received from forecast centers, local offices, however, have original forecast responsibility for issuing tornado warnings. Detailed and localized forecasts are prepared by these offices for their own localities. Although the localized forecasts are adaptations of the general area forecasts issued by the forecast center, they contain more detail and take into account the local climatology, topography, and weather peculiarities of the particular area to which they apply.

Local forecasts and reports are distributed by press, radio and television, public teletypewriter circuits, and automatic telephone forecast repeaters (in some large cities), and they are also available by telephone or personal visit at the Weather Bureau office During periods of severe weather, local Weather Bureau offices release emergency weather warnings to all radio and television stations in the area affected for frequent broadcast to the public. In addition, the Red Cross, Civil Defense, Coast Guard, state and local police, highway commissions, and other agencies are kept as fully informed as possible and assist in distributing warning messages so that every action may be taken to safeguard life and property [E.M V.]

Bibliography J Bjerknes and H. Solberg, Life cycle of cyclones and polar front theory of atmospheric circulation, *Geofysiske Publ.*, 3, 1922; T. F. Malone (ed). *Compendium of Meteorology*, 1951; J. Namias, Progress in objectivation and automa- [illegible]

[illegible] *Some Applications of Statistics to Meteorology*, 1958, S Pet-

terssen, *Weather Analysis and Forecasting,* 2d ed., 2 vols., 1956; N. A. Phillips, Recent Developments in Numerical Weather Prediction, in *Advances in Computers,* vol. 1, Academic Press, New York, 1959.

Weather map

A chart portraying the state of the atmospheric circulation and weather at a particular time over a wide area. It is derived from a careful analysis of simultaneous weather observations made at many observing points in the area. Such a chart gives the weather forecaster an integrated picture of the location, structure, and, when several successive charts are available, the motion and development of the various weather systems. From this study he may construct a prognostic chart, which portrays various weather features for selected times in the future.

Many kinds of weather maps are used, depending on the weather elements of immediate interest and their elevation above the ground. At a typical large weather analysis central, as many as 35 different charts are constructed for a given time Among the more common of these is the surface map (see Fig. 2), which portrays the weather at the earth's surface All the mapped weather elements (Fig. 1) except pressure are those directly observed at the weather station. Except over the oceans the so-called sea-level pressure is a fictitious quantity obtained by reducing the surface pressure to sea level by a special formula so that continuous isolines (called isobars) may be drawn through regions having the same pressure value. These isobars are important in portraying the winds, the physical structure of weather systems, and the location of fronts and air masses. *See* AIR MASS; FRONT.

The surface map, at least over ocean regions, represents a section through the atmosphere along an approximately horizontal surface. This procedure may also be used in constructing upper-level

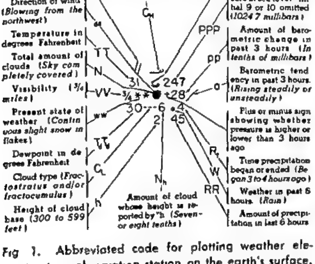


Fig 1. Abbreviated code for plotting weather elements at an observation station on the earth's surface. (U S Weather Bureau)

maps, or charts showing the distribution of weather at fixed elevations above sea level. It is more common, however, to portray the weather at high elevations on constant-pressure surfaces (Fig 3). The elements plotted at each upper-air station usually include the height of the constant-pressure surface temperature, and some measure of humidity, usually the dew point. Lines of constant elevation (called isohypses or contours) then are drawn to portray the topography of the selected pressure surface. The contours bear a relation to the winds similar to that of the isobars of a constant-level chart. Thus the winds tend to blow along the contours with low heights to the left when facing downstream. Their speed is roughly inversely proportional to the sine of the latitude and the distance between contours when these are drawn at constant height intervals. Usually isolines of temperature, or isotherms, are also drawn, as in Fig. 3

Other important two-dimensional or one-dimensional atmospheric sections cannot be termed weather maps, but rather meteorological charts and diagrams. These include vertical cross sections through the atmosphere, and thermodynamic diagrams similiar to those used in studies of heat engines.

Weather map analysis. This branch of synoptic meteorology had its beginning at about the time the telegraph was invented, when for the first time weather observations covering large areas could be sent rapidly to a central location. Its steady development since then was greatly accelerated following World War I when the techniques of air-mass analysis were developed by the Scandinavian school headed by V. Bjerknes. Upper-air charts were not commonly constructed until the 1930s, when sufficient high-level data first became available.

The preparation of a weather chart at a large analysis central can be described as follows First the encoded data at the surface and upper levels are transmitted to the central from collection centers by means of teletypes If a map covering the Northern Hemisphere is to be prepared, data from approximately 850 surface and 400 upper-air reporting stations must be processed This mass of data is subject to errors of observation, encoding, and transmission. Furthermore, large areas, particularly in the tropic and arctic latitudes, contain no observation stations and, hence, no data at all. The detection and correction of the errors, and the interpolation of weather in the intervals between reporting stations demand the greatest skill on the part of the chartmen who plot the data and the analysts who must interpret the data in terms of consistent physical structures of the atmosphere After the data are corrected and plotted, the analyst then locates and draws various features such as fronts, air masses, and isobars. He must be guided by known physical principles regarding the horizontal, vertical, and temporal continuity of the atmosphere, that is, his analysis must be internally consistent. When finished, the chart is ready for the forecaster.

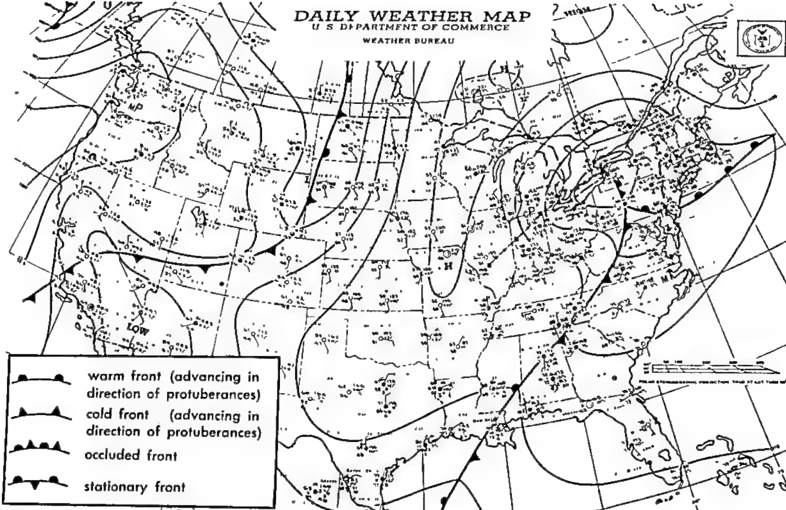


Fig 2. A surface weather map at 0600 Greenwich Civil Time, September 18 Sea-level isobars (thin lines) drawn for every 4 millibars and labeled in whole millibars, fronts, or transition zones separating air masses, indicated by heavy lines; *mT*, tropical maritime air; *mP*, polar maritime air. Areas where precipitation was falling at 0600 are shaded. Previous 6-hourly positions of low-pressure center in eastern Great Lakes indicated by crosses connected by arrows.

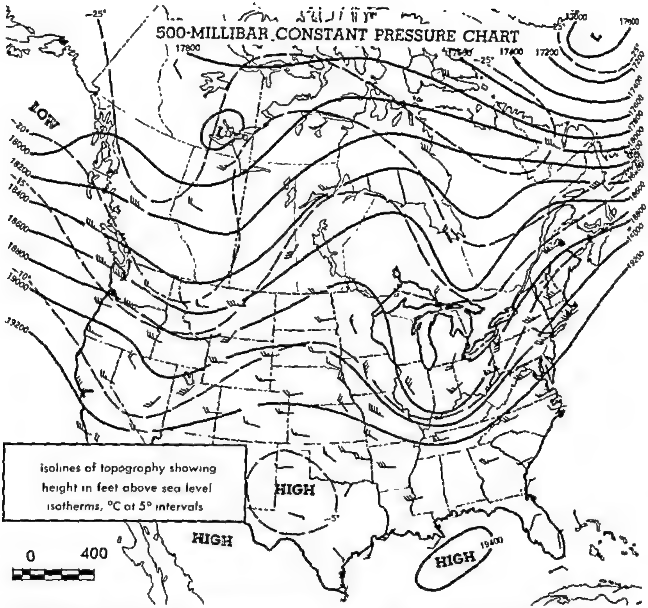


Fig. 3 Representative 500-mb constant-pressure chart. Arrows with barbs fly with wind and show wind speed in knots Each long barb equals 10 knots; triangular barb, 50 knots. (U.S Weather Bureau)

Automatic data processing. With the introduction of high-speed electronic computers into meteorology in 1951 it was immediately realized that the conventional method of data collection and processing described above was much too slow. A machine could produce a 24-hour prognostic chart for the entire Northern Hemisphere in less than an hour. It seemed hardly compatible that the collection and analysis of the raw data on which it fed should take almost 10 times as long Furthermore it was clear that short of setting up a huge weather factory employing hundreds of people it would become increasingly difficult to keep up with the larger volume of data pouring in from all parts of the world. For this reason experiments in automatic data processing and analysis were begun in a number of countries around 1953, and in 1957 the Joint Numerical Weather Prediction Unit at Suitland, Maryland, began the routine production of machine-analyzed weather maps. It is a tribute to the firm scientific foundation laid down by the earlier syn- reproduced in large part those used in conventional analyses While it has been found difficult to program the machine so as to duplicate precisely the technical skill of trained chartmen and analysts, the machine has two advantages, speed and precision

An obstacle to the ultimate goal in rapid data processing has been the painfully slow procedure for collecting and transmitting weather observations. Only thorough development of robot weather reporting stations, to feed data directly into the machines, can fully utilize the speed of modern computers. *See* DEW POINT; METEOROLOGY; WEATHER (FORECASTING AND PREDICTION). [P F C]

Weather modification

Modification of any aspect of the naturally occurring weather by the intervention of man. The usual purposes of weather modification are to reduce damage and hazards resulting from such phenomena as low temperatures, wind, fog, and drought

Modifications for agriculture. Agriculture is par-

[th]e important methods of weather control have [be]en used in this field

Frost protection. Protection against frost is [ne]eded in many areas Frost usually results from [th]e radiative cooling of the ground, favored by [cle]ar skies and light winds A relatively thin layer [of] air near the ground may cool below freezing [wh]ile the superincumbent air remains warmer.

Two common means of frost protection are (1) [el]evated blowers to mix the cold air with the [wa]rmer air aloft and (2) heaters on the ground. It [wa]s once thought that the heaters should produce [sm]oke to intercept the radiation from the ground, [bu]t it is now known that smoke is transparent to [su]ch radiation. Both methods fail when the low temperatures are due to the advection of air, which is generally cold at all heights and accompanied by strong winds

Other applications to agriculture. Deserving mention are irrigation, wind-breaks, soil coverings to reduce evaporation, and protective covers for plants. Partial success has been achieved in hastening the melting of snow to permit earlier planting by spreading a material which increases the absorption of solar radiation

Evaporation from water bodies. It has been found that the maintenance of monolayers of substances such as cetyl alcohol on reservoirs and ponds will reduce evaporation by 50% or more.

Fog control. Fog is a hazard to navigation, particularly for the landing of aircraft, and many methods of dispelling fog have been tried. Fog is an aerosol composed of drops, of the order of 10 microns radius, in equilibrium with the water vapor in the air Fog may be removed by evaporation or by the phys[illegible]

Heat me[illegible]
the princi[illegible]
operation)
War II [illegible]
[illegible] ensure prompt evaporation, the air must be heated to reduce the humidity; the heat required is several times the latent heat of vaporization. Concentrated heat sources cause convection, which brings in fresh fog Fido systems at airports burned fuel at rates considered excessive for commercial applications; the heat actually required exceeded the theoretical by many times because of convection

Physical methods. Success has also been achieved by releasing hygroscopic particles which absorb water vapor and permit the fog to evaporate This technique proved more efficient than Fido but is also expensive and involves cumbersome equipment and corrosive materials. In principle, fog may be removed by electrical precipitators or by dropping charged particles through it, but there are certain disadvantages and little work has been done Supercooled fogs may be dissipated by dry ice or silver iodide smoke using the principles discussed below.

Inducing rainfall. The artificial stimulation of rainfall has gained wide attention and some degree of success. The principal method is based on the ice crystal mechanism in which a few ice crystals form on natural freezing nuclei in supercooled clouds Because the equilibrium vapor pressure over ice is lower than over supercooled water, the ice crystals grow rapidly by diffusion and attain sufficient size to fall as precipitation particles There is evidence that the population of natural freezing nuclei is highly variable and that there are many occasions on which there is an insufficient number active at the prevailing cloud temperatures to form precipitation efficiently. Silver iodide, pellets of dry ice, and certain other substances act as efficient freezing nuclei at temperatures below about −5°C Seeding clouds with these substances can compensate for a deficiency of natural nuclei

Because of the complexity both of natural cloud systems and of the precipitation process, it is difficult to predict the results of seeding and recourse is usually had to a statistical analysis of the rainfall, comparing seeded with unseeded storms. Results indicate that the most favorable situations are those in which the clouds are formed by forced lifting over mountain ranges. In such cases, cloud water not precipitated evaporates on the lee side and the more rapid action of the artificial nuclei shows to best advantage. Results are not so favorable in large cyclonic storms over level terrain where ample time is available to the natural processes. Seeding can produce large holes in supercooled stratus decks. Warm clouds in which precipitation is formed by collision have been seeded with water drops and salt particles. *See* CLOUD PHYSICS.

Summary. Because of the tremendous energy of weather processes, forcible large-scale control by man appears unlikely A more sophisticated approach is required, based on a detailed understanding of natural processes. Such knowledge may provide a clue to control, probably based on a sensitive rearrangement of the energy sources and sinks. On the other hand, it may be found that the large-scale motions of the atmosphere are so stable that man cannot yet hope to influence them [H G.H.]

Weather station

A place and facility for the observation, measurement, and recording and transmission of data of the variable elements of weather. Although there are hundreds of private observers and forecasters in the United States, the most effective observations and forecasts are those made with the aid of the great network of the U S. Weather Bureau This article deals with the observation and data-gathering stations of this nationwide agency.

Types of stations. Most weather stations are classified as first-order or second-order largely on the basis of their primary or supplementary functions

First-order stations. Major weather stations located at airports measure temperature, dew point, wind, ceiling, and visibility over grassy plots near runways

Ceiling, the lowest height of clouds covering over half the sky, is measured by ceilometers. A beam of modulated light is projected on the cloud base. At a known distance, the spot of light is electronically

scanned day and night and the cloud height computed trignometrically. At smaller airports, the ceiling is measured during the day by timing the ascent to disappearance of a balloon rising at a known rate. At night, a nonmodulated light beam is used, and cloud heights are computed trignometrically.

Prevailing visibility is determined from markers at known distances during daylight or lights at known distances at night At jet terminals, transmissometers parallel to the instrumented runway are calibrated in units of runway visibility. At airports equipped with high-intensity runway lights, runway visual range, the distance a pilot can see high-intensity lights along the runway as he approaches for a landing, is reported This is determined by an electronic computer, using transmissivity, background illumination, and the intensity setting of the runway lights.

All observational data are remotely recorded in weather offices near the field. Rainfall intensity is remotely recorded, using tipping bucket, weighing, or float-type gages. Snowfall rates present numerous problems in telemetering. Pressure, pressure change, and altimeter settings are measured by mercurial or precision aneroid barometers and barographs. Determined by visual observations are clouds, current weather conditions, and obstructions to vision. Remote automatic stations transmit to a weather office by teletypewriter all observations of elements for which sensors are available.

Numerous airport and other first-order stations in larger cities measure solar radiation, ozone content of the air, and gradients of temperature and moisture at short intervals above and below the surface.

Many first-order stations are also equipped to make upper-air measurements. Precise theodolite observations of the course of a pilot balloon rising at a known rate provide data on the horizontal velocity of upper-level winds. A smaller number of stations measure pressure, temperature, and humidity by means of a radiosonde, carried by a larger balloon. Many stations use radio direction-finding or radar equipment to track radiosondes.

Second-order stations. Supplementary or second-order stations furnish detailed surface weather data from key points at specified hours Climatological stations manned by volunteers record the extremes of temperature and the amount and type of precipitation each day. Many of these stations also serve hydrology and in addition report snow density, river stage, rates of stream discharge, evaporation, and, in some cases, wind.

Over ocean and lake areas, reports are received from moving ships, specially equipped fixed ships, and a few automatic stations on anchored buoys.

Radiosonde data over oceans are limited to the fixed weather ships and a few moving ships. Only the fixed ships are equipped for upper-wind measurements. Important upper-air data over ocean areas are obtained by reconnaissance planes flying regular patterns but using alternate courses during storms Limited data over oceans can be obtained from transosondes. These large balloons, carrying radio-transmitting equipment, fly at a constant altitude and provide temperature and wind data In spite of the difficulties involved, both rockets and satellites must be used to obtain data from levels not otherwise probed.

Over land, at sea, and in the air, radar surveys all weather within a radius of 100–200 miles. Hydrologists can use radar photographs to interpolate between rainfall measurements and thus draw more complete isohyetal maps. For air safety, radar locates the active precipitation areas and areas of turbulence and, even when skies are clear, may identify important wind-shear lines.

Frequency of observations. Most first-order stations report to national networks of weather offices and centers at hourly intervals and to international networks at 6-hour intervals. Second-order stations report at 3- or 6-hour intervals. Upper winds are observed at 6- or 12-hour intervals and radiosondes are released every 12 hours. Moving ships report every 6 hours and fixed ships every 3. Hydrologic stations report daily during storms Climatological stations mail reports monthly. For hydrologic purposes, storage precipitation gages in mountainous areas are visited seasonally.

Density of observational networks. Spacing of observation points is strongly influenced by population density, with many observations near coasts, in valleys, and along main transportation routes. The required density of stations varies for each meteorological user, and there is no universally acceptable and economically realistic plan. However, meteorologists agree on broad objectives. Roughly, these are surface stations not more than 100 miles apart, upper-air stations not more than 300 miles apart, climatological stations 30 miles apart, and radar stations 200 miles apart. Most meteorologists would accept an ocean spacing which was one-fifth as dense as that over the continents. *See* METEOROLOGICAL INSTRUMENTATION. [A.K.S.]

Weathering processes

Weathering may be defined as the change of geologic materials (minerals and rocks) at or near the earth's surface from relatively massive to dispersed states. Climate, plants and animals, and such agents as water, wind, ice, gravity, temperature change, and the gases oxygen and carbon dioxide, play an important role in weathering.

Types of weathering. There are two general types of weathering: mechanical, in which rocks are disintegrated (broken into smaller fragments by physical forces), and chemical, in which rock materials are decomposed (changed in composition by chemical reactions). In cold and in dry climates the mechanical agents are most active and produce angular-shaped particles and landforms. In warm humid climates chemical reactions and biochemical changes proceed at much faster rates, particularly in the tropics where weathered zones sometimes extend to great depths.

Products of weathering. Products of mechanical weathering include rock fragments and materials formed by (1) the expansion and contraction in rocks, (2) the action of freezing water, and (3) the shattering effects of downslope movement. See EXFOLIATION, ROCK; FROST ACTION; MASS WASTING.

Products of chemical weathering include such economically and technologically important products as the soil; clays used in making structural products, ceramic whitewares, firebrick, portland cement, absorbents, catalysts, and fillers; and ores of iron, manganese, uranium, and vanadium. See CLAY, COMMERCIAL; ORE AND MINERAL DEPOSITS.

Weathered products are commonly designated as *relatively insoluble* (both those that remain in place and those carried away in suspension) and as *soluble* (those removed in solution by water). *Colloidal products* of weathering, however, obscure a sharp division between soluble and insoluble weathering products. The colloidal products are important both quantitatively and qualitatively.

Chemical weathering processes. During chemical weathering, chemical elements of geologic materials assume higher states of oxidation. Silicate rocks weather chiefly by hydrolysis, but also simultaneously by oxidation and carbonation, forming hydrates, carbonates, and oxides. For example, iron in silicate minerals combines with oxygen to form Fe_2O_3, thereby removing Fe from the silicate structure and so disrupting that network. The carbonation of calcium and magnesium in silicates also aids in their break-up. The general hydrolysis reaction of silicates may be written as follows:

$$MAlSiAlO_n + HOH \rightarrow M^+OH^- + HSiO_n$$
$$+ H(M)AlSiAlO_n + Al(OH)_3$$

M refers to metal cations (K, Na, Li), subscript *n* to an unspecified ratio of atoms, and the Al following Si substitutes for Si. Thus there are formed in hydrolysis soluble alkali-metal hydroxides, soluble silica (perhaps H_4SiO_4), and probably a relatively insoluble clay mineral (zeolite) or, less commonly, hydrated alumina. If the hydrolysis takes place at pH 9.5 or higher, both silica and alumina will be relatively soluble and mobile. They may then be separated and form bauxite ($Al_2O_3 \cdot nH_2O$). Under more acid conditions, clay minerals are formed.

Adding hydrogen ions to the hydrolyzing system increases the rate of reaction. Carbonic acid, formed when the carbon dioxide of the air and soil dissolves in water, is a source of hydrogen ions which accelerate the reaction. Calcium bicarbonate, $Ca(HCO_3)_2$, magnesium bicarbonate, $Mg(HCO_3)_2$, and carbonate-complexed compounds of various cations are removed in solution. Organic (humic) and other acids participate in the hydrolysis. Another major source of hydrogen ions is their production in the ionic atmosphere about the rootlets of growing plants. During plant growth and metabolism, hydrogen ions are evolved; these are exchanged by the roots for nutrient cations (K^+, Ca^{++}, Mg^{++}) present in nearby clay colloids and rocks. Thus the process of nutrition for plants is simultaneously a process of weathering for rocks. Plants that are primitive in development apparently possess higher energies of cation exchange than do those more advanced; lichens derive nutrient cations from fresh rock without intermediary soil. The growth of plants is therefore exceedingly important in weathering. Chelating organic substances extract cations from rocks, implementing rock breakdown. Partial weathering makes the rock constituents more available to plants, but extended weathering removes the nutrient materials entirely.

The ultimate destination of the soluble products of weathering is the ocean. There the dissolved mineral matter becomes concentrated in solution and in deposits. Potassium, although as soluble as sodium, is more readily absorbed by clay minerals as exchangeable cations and may be incorporated in the crystals of hydrous mica. Potassium is, therefore, less concentrated than sodium in sea water. Magnesium may be incorporated in chloritic varieties of clay minerals. See CLAY MINERALS.

The most abundant weathering products of silicate rocks are the clay minerals. Weathering (hydrolysis) resulting in high concentrations of calcium, magnesium, and iron (ferrous) ions tends to form the montmorillonite group of clays. Such high concentration of ions occurs where evaporation exceeds precipitation, ground-water drainage is poor, or hydrolysis is rapid (as in weathering of volcanic dust). The kaolin group of clay minerals is developed where rainfall exceeds evaporation and leach-

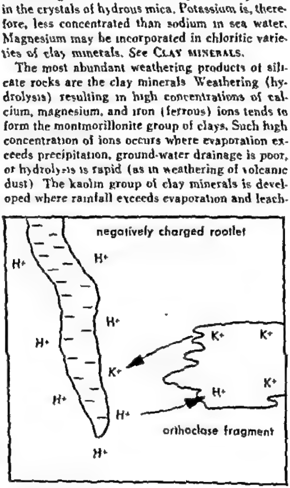


Diagram showing conversion of orthoclase to a clay mineral by plant roots. A hydrogen ion (positive) from those surrounding the negatively charged rootlet replaces a potassium ion in the orthoclase fragment. The bonding of the hydrogen ion with the oxygen in the original mineral begins the conversion of the orthoclase to clay. The replaced potassium ion replaces the hydrogen ion along the rootlet and is eventually utilized in plant growth. (Adapted from W. D. Keller and A. F. Frederickson, Role of plants and colloidal acids in the mechanism of weathering, Am. J. Sci., 250:603, 1952)

ing is intense. Oxidation of iron is then ordinarily high. Under conditions of very drastic leaching and continual wetting of the rocks, as in a tropical rainforest, silica and most of the cations are dissolved leaving hydrated oxides of alumina and ferric iron (bauxite and laterite).

Limestones and dolomites weather chemically by dissolution of the calcium and magnesium carbonates, leaving behind the less soluble impurities: quartz (sand), chert, iron oxides, and clay minerals. Clay minerals are subject also to breakdown during weathering by the removal of (1) exchangeable cations, (2) the more tightly fixed potassium of illite (hydrous mica) and possibly cations (other than Al) in the octahedral layer, and (3) silica. The clay minerals whose crystal structure is partly destroyed are said to be degraded. Entirely desilicated clays become bauxite or laterite. *See* BAUXITE; LATERITE. [W.D.K.]

Bibliography: W. D. Keller, *Principles of Chemical Weathering*, 1957; B. Mason, *Principles of Geochemistry*, 2d ed., 1958; P. Reiche, *Survey of Weathering Processes and Products*, Univ. New Mex. Publ. Geol. Ser., no. 1, 1945.

Weber

The unit of magnetic flux, also the unit of magnetic pole strength, in the meter-kilogram-second (mks) system of units.

In representing magnetic induction B by lines drawn in the direction of B, a number of lines is selected so that the number per unit area of a surface perpendicular to the magnetic induction is equal to B (*see* INDUCTION, MAGNETIC; MAGNETIC FLUX). When B is measured in newtons/ampere-meter and the area is in square meters, the flux representing the magnetic induction is measured in webers. A weber is one line of induction, as here selected. Flux density or magnetic induction is then measured in webers per square meter. Since the weber per square meter is equivalent to a newton/ampere-meter, a weber is a newton-meter/ampere. *See* ELECTRICAL UNITS; MAGNETIC FIELD; MAGNETOSTATICS. [K.V.M.]

Wedge

A piece of resistant material whose two major surfaces make an acute angle. It is closely related to the inclined plane and is used to multiply the applied force and to change the direction in which it acts (Fig. 1).

Force F is the smaller applied force and Q is the larger force to be exerted. In the absence of friction, forces must act normal to their surfaces; thus the actual force on the inclined surface is not Q but a larger force F_n. Summing up forces in the horizontal and vertical directions, shows that

$$F_n \sin\theta - F = 0$$

and

$$Q - F_n \cos\theta = 0$$

Combining the expressions for F and Q and solving for F gives

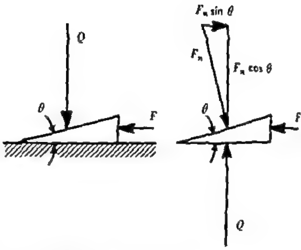


Fig. 1. Forces acting on a wedge.

$$F = Q\tan\theta$$

If angle θ is small, the reaction of Q against F is exceeded by the friction between the face of the wedge and the adjacent body on which it rests. Thus the wedge tends to remain in position even when loaded by a large force Q.

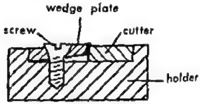


Fig. 2. Wedge plate used as a clamp to retain cutter in holder.

Some applications of the wedge are in splitting wood, in raising the platform of low-lift platform trucks, in cone clutches, and in combination with friction to fasten parts together (Fig. 2). *See* SIMPLE MACHINE. [R M.PH.]

Weevil

The common name for any species of coleopteran insect which can be distinguished from other beetles by the prolongation of their mouthparts into a snout or beak. Actually, both larval stage and adult are destructive for many plants of economic importance. The adult female deposits her eggs on or within plants. The larvae tunnel through the plant tissues, the adults attack the foliage. Both larval and adult stages are phytophagous and almost every part of a plant may be attacked.

The weevils include many of man's most serious insect enemies. The boll weevil, *Anthonomus grandis*, still destroys the equivalent of 1,500,000 bales of cotton each year in the United States, and the amount spent in its control is tremendous. More than any other force, this insect was responsible for crop diversification in the South.

The white-fringed beetle, *Graphognathus leucoloma*, is another serious weevil pest of the South, especially interesting in that it is wholly parthe-

genetic. It has rudimentary wings and is flightless.

Other weevils of economic importance include the plum curculio, which attacks most of the pit fruits and apples; the white pine weevil, the larvae of which develop in the bark of the pine twigs, causing great damage to the new growth; alfalfa weevils, the young of which develop in the alfalfa stems; and a large variety of weevils destructive to stored foods, especially beans, peas, and cereals, the young developing within the seed. *See* COLEOPTERA. [J D B]

Weight

The weight of a body is the force with which the earth attracts the body. By extension, the term is also used for the attraction of the sun or a planet [illegible]

[illegible] earth decreases at higher latitudes, and because the centrifugal force of the earth's rotation is greatest at the Equator, the weight of a body is smallest at the [illegible]

[illegible] the earth, it diminishes with altitude and with the depth below the surface. For example, the weight of a body diminishes by about 0.1% if it is raised 2 miles above the earth's surface or taken 4 miles below the surface. Weight also depends to a smaller but measurable degree on the density of the earth's crust below the body. Weight is measured by several procedures. *See* BALANCE (WEIGHING INSTRUMENT), WEIGHT MEASUREMENT.

Since weight is a force, it is expressed in force units. In the United States, the commonest unit of weight is the pound, sometimes written pound force or pound weight, to distinguish it from the mass unit, pound (*see* POUND). Pound weight is the weight of a 1 lb mass at a location where the acceleration of gravity is 32.174 ft/sec². Where the acceleration of gravity is g, the weight of a 1 lb mass is $g/32.174$ lb weight. One grain is defined as 1/7000 of the pound weight just defined. Besides this pound, legally known as pound avoirdupois, there is pound Troy, which is equal to 5760 grains, and the Apothecaries' pound, which is also equal to 5760/7000 lb avoirdupois. Units intermediate between the pound and the grain are for pound avoirdupois, ounces and drams; for pound Troy, ounces and pennyweights, and for the Apothecaries' pound, ounces, drams, and scruples. Only the grain denotes the same weight in all three systems. A carat, in terms of which the weight of precious stones is often given, is really a unit of mass, equal to 0.2 g. *See* MEASURE. [L.N.]

Weight measurement

Weight is the resultant force acting on a mass (in a vacuum) due to the earth's gravitational field. Units of weight are based upon an acceleration of gravity of 980.665 cm/sec² or 32.1740 ft/sec². When the weight of an unknown is determined by comparison with a known weight, there is no error in the readings due to gravity variations. The varying buoyant effect of the atmosphere is negligible when the density of the unknown is approximately the same as that of the standard. In precision weighing, this buoyant effect of air must be considered.

Weight, unlike temperature, pressure, liquid level, and similar variables, is directly related to mass, a basic characteristic of matter. It is widely used as a measure in the transfer of materials, in synthesis and analysis, and in blending and separation. Therefore, it is frequently desirable to measure weight to a greater accuracy than most variables. Weighing devices are designed for a maximum load and the specified accuracy to operate within a given temperature range. They must be rugged enough to withstand service conditions, which may be corrosive and abusive.

Weighing is a dynamic operation, and the time required for a scale or balance to reach an equilibrium condition is important. Mechanical and spring balances are often equipped with fluid dashpots or other devices to critically damp the system, permitting equilibrium to be reached in the minimum time. The more elaborate systems (electronic and pneumatic) have inherently short time constants and therefore reach equilibrium quickly. The pneumatic system involves feedback and therefore requires dynamic consideration in the design of the load element to obtain stability. Most weighing devices may be used directly, or with minor modifications, for the measurement of other forces.

Simple balance. This device consists of a lever and pans supported on a knife edge or flexure (see Fig. 1). Available in many forms, it is used to compare known and unknown weights. The accuracy, sensitivity, damping characteristics, and load capacity of a balance are determined by its design and physical condition. The same equal-lever-arm principle is used for precision laboratory balances

Fig. 1 Simple balance (Torsion Bal[illegible]

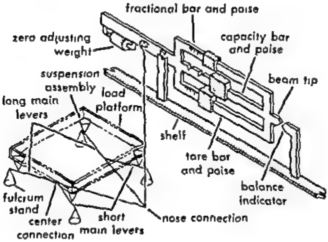


Fig. 2. Mechanical-type industrial scales. (*From D. M Considine, ed , Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

which have a sensitivity as high as 0.01 mg (weigh to 1 part in 1,000,000 or better). Most commercial and industrial balances of this type are not designed to weigh to better than 1 part in 10,000. Note that the final balancing is accomplished by adjusting a small fixed weight (slider) on a lever arm.

Mechanical-type industrial scale. This type incorporates a number of levers with precisely located fulcrums to permit heavy objects to be balanced (weighed) with small, convenient counterweights or counterpoises (see Fig. 2) Shaft pivots, *knife edges, cone pivots, and flexures of various* types are used at the fulcrum points and are designed to carry maximum load at that point in the lever system with minimum friction. Multiple-lever systems support the platform, or hopper, so that unequally distributed loads can be accurately weighed. Often the counterweights on an industrial scale are separated for the convenience of the operator. The zero-adjusting weight balances the scale with no load on the platform. The tare counterweight balances the scale carrying an empty container. The load is balanced most often by counterweights of fixed mass at a definite location in the multiple-lever system plus at least one counterweight (poise) which can be moved to increase or reduce its distance from the final pivot. Truck and railway scales frequently use this same principle. While the accuracy of scales of this type varies with the design and physical condition, the commercial accuracy tolerance is 0.1% of the maximum capacity, and the sensitivity of the scale is 0.05%. Somewhat higher accuracy and sensitivity can be achieved by refined designs.

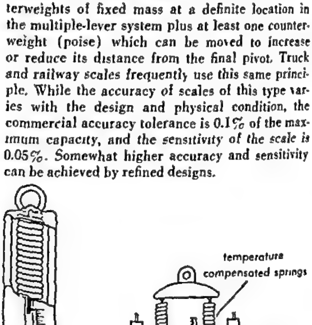


Fig 4 Spring scales. (*a*) Straight-faced. (*b*) Dial-type (*From D. M Considine, ed , Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

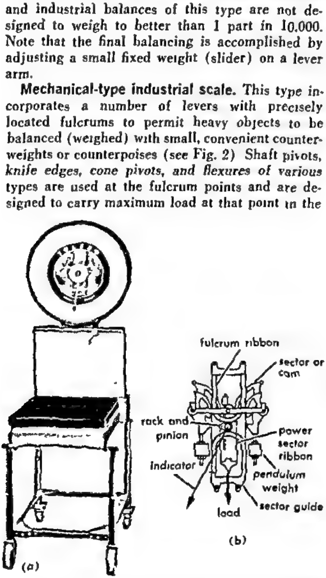


Fig. 3. Pendulum-type mechanical scale. (*a*) Typical commercial model (*Toledo Scale Co*). (*b*) Schematic detail at pendulum (*from D. M. Considine, ed , Process Instruments and Controls Handbook, McGraw-Hill, 1957*).

Pendulum-type mechanical scale. This type balances the force of the load by the rotation of a bent lever (Fig 3). With this construction, the deflection of the load on the scale moves the counterweights through the lever system so that their center of gravity is at a greater distance from the final fulcrum. Thus the increased lever arm of the counterweights automatically balances the load. Normally this movement, through a *rack and pinion*, moves a pointer on a calibrated dial. Cams are often incorporated in the lever system to linearize the movement of the pointer on the dial. Commercial scales of this type are also accurate to 1 part in 1000

Spring scale. This type utilizes the deflection of a spring to measure the load Two basic designs are illustrated in Fig. 4.

Hydraulic systems. In hydraulic systems, such as those in Figs. 5 and 6, the load cell carries the force by a hydraulic pressure acting upon a cylin-

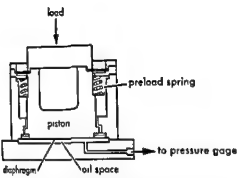


Fig 5 Hydraulic scale using load cell (*From D. M. Considine, ed, Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

der or capsule of known effective area The pressure is measured at a remote point by a pressure gage, such as a Bourdon tube (*see* PRESSURE MEASUREMENT). Some load cells convert all the force to pressure by means of a slack diaphragm (Fig 5), which is accurate to 1 part in 400; others carry the bulk of the load directly, using the pressure primarily for transmission purposes (Fig. 6), and are accurate to 1%. Hydraulic weighing systems are temperature-sensitive, and while they may be compensated, it is customary to provide a tolerance for temperature variations.

Pneumatic systems. Pneumatic systems, such as in Fig 7, detect the load by a sensitive nozzle and flapper system and balance the load by modulating an air pressure in an opposing capsule. The effective area of the capsule must remain constant; this is accomplished by minimizing the motion of the platform so that the system approaches a null balance. Maximum errors as low as 0.25% are possible at a constant operating temperature.

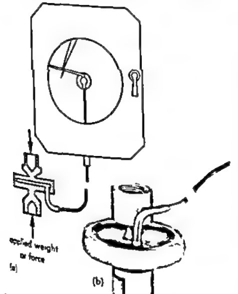


Fig 6. Volumetric load element (a) Hydraulic system using volumetric load cell (b) Commercial volumetric load element (*Taylor Instrument Co*)

Electrical weighing systems. These systems, such as in Fig 8, usually involve the electrical measurement of the elastic deformation of a mechanical element under stress. The strain gage is attached to the weighing element in a manner to produce the maximum resistance changes per unit of load. The change in resistance with load is measured and amplified by electronic means, and the load is read on a potentiometer. Usually the elements of the strain gage are the arms of an ac bridge circuit, and the circuit is carefully designed to minimize ambient-temperature and supply-voltage effects, as well as zero drift Normal commercial tolerance on this type of weighing system is 0.25% at a given temperature Better accuracy can be achieved through special design *See* STRAIN GAGE.

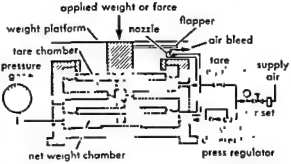


Fig 7 Pneumatic scale (*From D M Considine, ed, Process Instruments and Controls Handbook, McGraw-Hill, 1957*)

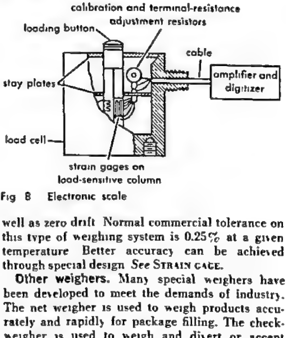


Fig 8 Electronic scale

Other weighers. Many special weighers have been developed to meet the demands of industry. The net weigher is used to weigh products accurately and rapidly for package filling. The checkweigher is used to weigh and divert or accept packages which have been filled The continuous strip weigher and the continuous product feeder provide continuous measurement of a moving product. Such weighers must have good dynamic characteristics (low mass, high stiffness, short time constant) The maximum error of these weighing devices varies between 1% and 0.25%.

Scales may be of the indicating type (dial or digit), of the recording type (chart or printed tape), or may be equipped with controlling features such as cutoffs, continuous automatic feed-

ers, and alarms. Scales are also classified by their use, for example, druggists' scales, crane scales, platform scales, or railway scales. Scales are also available for many additional specialized services, such as counting by weight or inspecting by weight. [R.E CL.]

Weightlessness

Strictly defined, weightlessness means complete and total absence of weight or zero gravity. This word is loosely and commonly used to denote subgravity or a gravitational force of less than the normal 1 *g*. *See* GRAVITATION.

In space flight, a state of weightlessness exists when the gravitational pull of the earth or other planets is at zero. In such a state, objects and subjects tend to float. Experimental evidence shows that under such conditions, individuals, because of the loss in position sense (proprioceptive deprivation), tend to become disoriented and confused, but can adjust as long as tactile and visual references are intact (*see* KINESTHETIC SENSATION; PERCEPTION). Apprehension and anxiety are common under such conditions. There is no evidence of impaired locomotion or muscular control. However, cardiovascular and respiratory changes have been recorded These latter changes are probably secondary to emotions and to the hemodynamics of acceleration and deceleration rather than anything related to weightlessness per se. Under conditions of zero gravity, it is impossible to eat and drink unless water and food are placed sufficiently far back in the mouth so that the normal reflex of deglutition can propel the water and food down the esophagus. *All in all, weightlessness itself should not* produce any untoward or ill effects or impose any insurmountable difficulty to space flight. *See* SPACE BIOLOGY. [M.CO]

Weil's disease

An acute febrile disease of man produced by spirochetes of many species of *Leptospira*. The disease is also called leptospirosis. The illness follows a general pattern However, even cases due to a single species exhibit all degrees of severity; some infections are so mild that they are recognizable only by serological tests. The incubation period is 6–15 days. Among the prominent features of the disease are fever, muscle pains, headaches, hepatitis, albuminuria, and multiple small hemorrhages in the conjunctiva or skin. Meningeal involvement often occurs. The febrile illness subsides after 3 to 10 days. Fatal cases show hemorrhagic lesions in the kidney, liver, skin, muscles, and central nervous system

Jaundice, which is a prominent feature of Weil's disease, originally believed to be due only to *L. icterohaemorrhagiae*, is known also to be caused by other species. *L. icterohaemorrhagiae* and *L. canicola* tend to give rise to a more severe type of human disease than do *L. hebdomadis*, *L. pomona*, *L. grippotyphosa*, *L. automnalis*, *L. australis*, *L. mitis*, and *L. batavia*. Definite diagnosis is by demonstration of leptospira in the blood or spec antibodies in the serum. Early in the disease, organisms may be isolated from the blood by ture or by intraperitoneal inoculation of yo guinea pigs or hamsters. Serum antibodies to specific type of leptospira develop and rise shar from the seventh to the fourteenth day and demonstrated by agglutination of living organi or by complement fixation tests employing extra of leptospira as antigens Antibodies persist many years. *See* ANTIBODY.

No specific treatment yields dramatically ef tive results. Tetracyclines are the preferred dru

Wild rodents are the principal reservoirs, though natural infection occurs in swine, cat horses, and dogs and may be transmitted to m through these animals. Man is infected eit through contact with the urine or flesh of disea animals, or indirectly by way of contamina water or soil, the organisms entering the bo through small breaks in the skin or mucous me branes. The incidence of leptospirosis is high among sewer workers, slaughter-house employe workers in rice and cane fields, and bathers stagnant pools. Preventive measures are direct towards elimination of rat infestation and the we ing of protective clothing. No effective vaccine available. *See* SPIROCHETE. [T.B]

Welded joint

The joining of two or more metallic components introducing fused metal (welding rod) into a fil between the components or by raising the tempe ture of their surfaces or edges to the fusion te *perature and applying pressure (flash welding)*.

Types of welded joint. In a lap weld, the edg of a plate are lapped one over the other and t edge of one is welded to the surface of the oth (Fig. 1).

In a butt weld, the edge of one plate is brought line with the edge of a second plate and the joi is filled with welding metal or the two edges a resistance heated and pressed together to fuse.

For a fillet weld, the edge of one plate is broug against the surface of another not in the same pla and welding metal is fused in the corner betwe the two plates, thus forming a fillet. The joint ca be welded on one or both sides

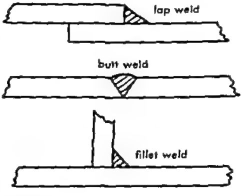


Fig 1. Welded joints

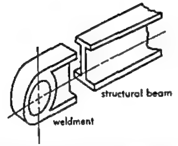


Fig 2 Weldment is a preformed component designed to be welded to other structural or machine components.

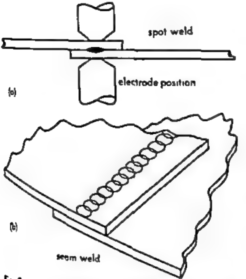


Fig. 3. Common electrical resistance welds. (a) Spot weld. (b) Seam weld.

A weldment is a cast steel, forged steel, or machined steel component that is assembled to plates or structural steel shapes or other weldments by welding to form a machine part (Fig 2).

A spot weld is an electrical resistance lap weld wherein lapped surfaces are brought to high temperature by the resistance to a low-voltage, high-amperage current between a pair of water-cooled electrodes (Fig. 3). A spot weld is used primarily for thin sheet stock

A seam weld is similar in production to a spot weld except that the electrodes are rollers or wheels rotating at constant speed; they produce a long narrow weld The butted edges of pipe or tubing can be welded by a modified seam weld

Strength of welded joints. Because welded ments are usually exposed to a complex stress pattern as a result of the high temperature gradients present when the weld is made, it is customary to design joints by use of arbitrary and simplified equations and generous safety factors

The force F of direct loading and consequently the stress S is applied directly along or across a weld. The stress force equation is then simply $F = SA$ in which A is the area of the plane of failure (Fig. 4).

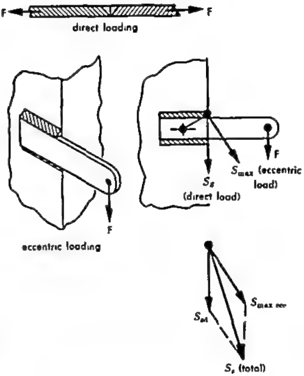


Fig 4 Loading forces on a welded joint.

For eccentric loading, the force F causes longitudinal and transverse forces of varying magnitudes along the weld. The stress is found by assuming that rotation occurs around the centroid of the welded area and that the shear stress due to the torque about the centroid is vectorially added to the direct shear caused by the force Several points in the welded area must be checked to establish the maximum vectorial summation of the direct and eccentric stress Maximum shear stress S_{max} is $Fd\, r_{max}/J_{total}$ where J_{total} is total polar moment of inertia of weld failure sections about the centroid. d is the distance from force F to the centroid, and r_{max} is the radius from the centroid to the farthest part of the weld. See RIVETED JOINT; STRUCTURAL CONNECTIONS, WELDING AND CUTTING OF METALS. [L.S L.]

Bibliography· ASME, *Boiler Code*, 1958; T. B Jefferson and W. J Brooking, *Introduction to Mechanical Design*, 1951, V L. Maleev and J. B. Hartman, *Machine Design*, 3d ed , 1954; A. Vallance and V. L. Doughtie, *Design of Machine Members*, 3d ed , 1951

Welding and cutting of metals

Processes based on heat to join and sever metals Welding and cutting are grouped together, because in many manufacturing operations severing precedes welding and involves the same production personnel Welding is one of the joining processes, others being riveting, bolting, and gluing The American Welding Society's definition of welding is "a metal-joining process wherein coalescence is produced by heating to suitable temperatures with or without the application of pressure, and with or

without the use of filler metal." Brazing is defined as "a group of welding processes wherein coalescence is produced by heating to suitable temperatures above 800°F and by using a non-ferrous filler metal having a melting point below that of the metals to be joined. The filler metal is distributed between the closely fitted surfaces of the joint by capillary attraction." Soldering is similar in principle, except that the melting point of solder is below 800°F. The adhesion of solder depends not so much on alloying as on its keying into small irregularities in the surfaces to be joined. For comparison of metal joints *see* JOINT (MECHANICAL).

Cutting is one of the severing and materials-shaping processes, some others being sawing, drilling, and planing. Thermal cutting is defined as a group of cutting processes wherein severing is effected by melting or by the chemical reaction of oxygen with the metal at elevated temperatures.

Welding and cutting are widely used in building ships, machinery, boilers, structures, atomic reactors, aircraft, railroad cars, missiles, and pressure vessels, as well as in constructing piping and storage tanks of steel, stainless steel, aluminum, nickel, copper, lead, and their alloys. For many products welding is the only joining process that achieves desired economy and properties, particularly leaktightness

Welding processes. Nearly all industrial welding involves fusion The edges or surfaces to be welded are brought to the molten state The liquid metal bridges the gap between the parts. After the source of welding heat has been removed, the liquid solidifies, thus joining or welding the parts together. The three principal sources of heat for fusion welding are electric arc, electric resistance, and flame. Hence arise the terms for the three fusion welding processes: arc welding, sistance welding, and gas welding. The flame in g welding is provided by the combustion of a fi gas.

Arc welding. The greatest volume of welding done with arc-welding processes. Arc welding w consumable electrodes is more widely practic than welding with nonconsumable electrodes consumable electrode is melted continuously by arc, one pole of which is the electrode, the oth pole being the metal to be welded. Generally, consumable electrode is a wire of the same chem cal composition as the work. The arc melts electrode and some of the base metal to form a p of weld metal, which after freezing becomes weld A nonconsumable electrode is tungsten carbon, which have high melting points and hi electron emissivities and therefore are not co sumed, except slowly by vaporization. (See Fig. 1

A covered electrode consists of a rod of me covered with material serving electrical and met lurgical purposes. Electrically, the covering in lates the rod from accidental contact with adjace material during welding, and provides an arc fr from interruptions. Metallurgically, the coveri may provide gas- and slag-forming ingredients protect the weld from the air, and it may supp deoxidizers or alloying elements to produce sou welds having specified chemical composition the arc consumes the electrode, the operator ma ually feeds it into the weld.

About 500,000,000 lb of covered electrodes a manufactured every year with diameters of ¹⁄₁₆- in. in steel, stainless steel, and nonferrous allo The weight of the covering may be from 10–50 of the weight of the covered electrode For weldi mild steel, covered electrodes are made in a nu

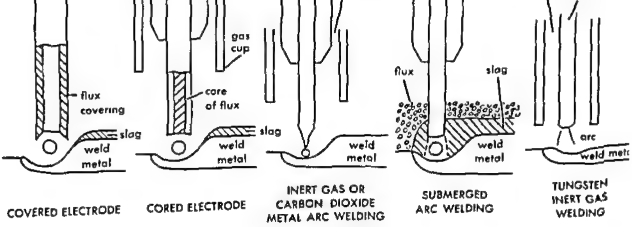


Fig 1. Five major arc-welding processes

ber of types, which have been classified by the American Welding Society. Class E-6010 refers to a covering consisting principally of cellulose The electrode can be used readily to weld flat, vertical, and overhead surfaces. Class E-6012 electrodes have a thin covering high in rutile (titanium dioxide) and are adapted to welding across gaps up to ¼ in. Class E-6016 electrodes are the low-hydrogen class The calcium carbonate in the covering evolves carbon dioxide during welding. The covering is nearly free from hydrogen, and the weld therefore is free from defects due to hydrogen Class E-6024 electrodes have about 40% iron powder in the covering, which makes possible high rates of depositing weld metal. The rate of depositing steel weld metal from steel electrodes in any arc welding process ranges from 0 03 to 0 06 lb per min per 100 amp of welding current.

Cored electrodes consist of a tube formed from strip and filled with *slag-forming*, arc-stabilizing, and alloying materials. For welding mild steel, the cored electrode is continuous In automatic welding the arc is guided mechanically along the joint, whereas in semiautomatic welding, the operator guides the arc manually by means of a torch In some processes, the arc is in air. In others, a gas, such as carbon dioxide, protects the arc from the air

Submerged arc welding is performed with a continuous electrode and a granular flux composed of silicates with or without deoxidizers and alloying elements. The flux is piled to a depth of ½–2 in along the joint. The arc melts some of the flux and is submerged in the liquid slag so produced. While submerged arc welding often uses currents of about 800 amp it can use as much as 4000 amp, far above that usable in any other arc-welding processes. High currents enable the weld to penetrate more deeply below the surface of the work than with other arc welding processes.

Inert gas metal-arc welding with consumable, continuous electrodes (called MIG welding, for Metal Inert Gas) requires no flux and produces welds without a slag cover. The arc is in an atmosphere of argon or helium supplied at 5–100 ft³/hr from the gas cup of the torch. Inert gas shielding is particularly advantageous in welding reactive metals such as titanium, which react with siliceous fluxes, and in welding metals such as aluminum and stainless steel, the corrosion resistance of which suffers in the presence of flux residues.

Carbon dioxide welding with continuous electrodes is similar to inert-gas metal-arc welding, except that CO_2 is used as a low-cost shielding gas. The process is restricted to steel. To prevent porosity resulting from the formation of carbon monoxide, CO, by the reaction of the carbon dioxide with the carbon in the steel, electrodes for CO_2 welding contain deoxidizers: manganese, silicon, aluminum, titanium, or zirconium.

Nonconsumable arc-welding electrodes are not deposited as part of the weld metal The electrode is one pole of the arc, usually the electron-emitting (and therefore cool) negative pole. The other pole is the work, which is melted. If additional metal is required to fill the joint, a filler wire is fed into the arc. Only tungsten and carbon have sufficiently high melting points to provide requisite electron emission at the high currents used in arc welding. Carbon electrodes seldom are used for welding because they vaporize more rapidly than tungsten.

Inert gas shielding is essential with tungsten electrodes, hence the term Tungsten Inert Gas (TIG) welding. A common electrode diameter is ⅛ in. using 150 amp. The process is adapted to welding thin material, ⅛ in and less, and the root, or first, pass in pipes, because the operator can control penetration more readily than with most other arc welding processes Direct current (electrode negative) is used in welding ferrous materials, copper, and nickel alloys. For aluminum and magnesium alternating current is required, because oxide film on the work is removed by reverse polarity (electrode positive). To prevent extinction of the arc as the electric potential passes through zero, a small high-frequency current (100,000 cycles) must be superimposed on the 60-cycle welding current

For a discussion of arc-welding equipment *see* ARC WELDING

Resistance welding. Resistance-welding processes are widely used in the manufacture of sheet metal assemblies, such as automobile bodies, and in joining bars and rods. By definition, the coalescence of the parts to be joined is accomplished by pressure applied to surfaces heated by resistance to electric current flowing in a circuit of which the work is a part In the most widely used processes, spot, seam, and flash welding, the heat melts the surfaces to be welded Every resistance weld involves a sequence of electrical energy and mechanical pressure. The sequence is provided by a control, which governs timing of both For example, a control meters the 10 cycles of 60-cycle power at 10,000 amp and also the electrode force at 500 lb required for spot welding two sheets of 0.040 in. steel The welding operator merely presses the button that sets the control in operation (See Fig 2)

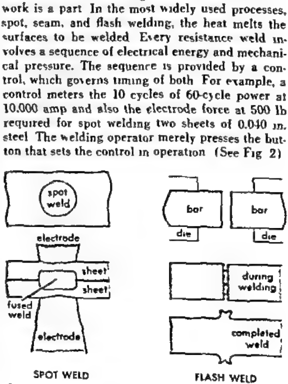


Fig 2 Spot and flash welding

In spot welding, the sheets to be welded are held between electrodes of hard, high-conductivity copper alloy. They conduct heat away from the electrode-to-sheet contacts. The contact resistance heat at the sheet-to-sheet contact fuses the sheets together. Electrode pressure provides uniform sheet-to-sheet resistance, and prevents expulsion of metal between the sheets

Seam welding is a process for making overlapping spot welds by means of rotating electrode wheels. The joints can be made gastight and liquid-tight, as in tubing and stainless steel electric refrigerators

In flash welding the surfaces to be welded are held lightly in contact while a high current flows through the few contact points. Melting occurs instantaneously at these "bridges." Violent vaporization ensues, with expulsion of small particles of hot metal as visible flash. Flashing is continued until the surfaces are coated with a layer of molten metal. They are then squeezed together to form the weld.

For further information *see* FLASH WELDING; RESISTANCE WELDING; SPOT WELDING.

Gas welding. A gas flame is a less-concentrated and lower-temperature source of heat than an arc. For this reason, a welding torch is often used for brazing, for welding thin material, and for applications requiring low gradients of temperature to avoid cracking, as in welding cast iron. The oxyacetylene flame, using a mixture of equal volumes of oxygen and acetylene (C_2H_2), has a higher temperature than any other commercially available fuel gas combination. High flame temperature is essential for the rapid localized heating required in welding Several sizes and types of torch tips are available to secure flames of different intensities and dispersions The ratio of oxygen to acetylene governs the chemical action of the flame. If the ratio is higher than unity, for example 1 1:1, the flame is oxidizing and is useful for preventing hydrogen porosity in brass. With a ratio below unity, 0 95:1, an acetylene feather appears in the flame, carbon is picked up by molten steel, and mill scale on the surface of steel is reduced A neutral flame with a ratio of 1:1 is used for steel. Flux is not used in welding steel, but is necessary for gas welding nonferrous metals. The use of other fuel gases, such as hydrogen, is restricted to the welding of metals of low melting point, such as lead.

Other fusion-welding processes Among other fusion-welding processes are the induction, thermit, and hot air processes. The heat for induction welding is obtained from the resistance of the work to induced high-frequency current. Proximity effect concentrates the current in the edges to be joined, as in the manufacture of metal radio tubes. In ther- [illegible] ylene, are welded by a [illegible] current of hot air melting the edges that are to be joined.

Nonfusion processes. In these processes, the surfaces to be welded coalesce in the solid state by application of pressure with or without heat. It is essential that the surfaces be free from films, such as oxides. Resistance upset welding uses contact resistance heat to bring the surface to a high temperature but below the melting point. Unlike flash welding, the surfaces are squeezed tightly together throughout the process, which upsets, or changes the dimensions of, the metal close to the weld In forge welding, the parts are heated in a furnace and pressure is applied usually by hammer blows The heat for gas-pressure and induction-pressure welding is derived from flames and induced electric currents, respectively. Cold welding is performed by pressure alone without heat, and is applicable particularly to soft metals, such as aluminum. Ultrasonic welding also requires no heat and only light pressure; the atomic movement required for coalescence is stimulated by ultrasonic vibrations

Welding metallurgy. The effects of welding on a product often govern its suitability for service The weld metal, the zone close to the weld (called the heat-affected zone), and the thermal stresses and strains caused by welding may combine to yield a product different in strength, ductility, leaktightness, and dimensional stability than the unwelded base metal Weld metal frequently is compared with cast metal with which it shares the characteristic of freezing in columnar dendritic pattern in the direction opposite to the direction of withdrawal of heat. Unlike cast metal, weld metal usually has a free surface exposed to the air, and it is free from shrinkage cavities. Weld metal freezes rapidly compared with most castings and is comparatively fine-grained and free from chemical segregation. Porosity must be guarded against It is a result of evolution of gas, such as nitrogen from steel, during freezing of the gas-saturated weld pool Porosity also may result from a chemical reaction in the weld metal, such as the reaction between carbon and oxygen in liquid steel Hot cracking of the weld metal during the last stages of solidification may be prevented either by fixturing so as to avoid strain during freezing, or by adding suitable alloying elements to the weld, as in the addition of manganese to steel weld metal to combine with sulfur.

The heat-affected zone may be softer than the base metal, especially in alloys that have been cold-rolled or heat-treated for high strength. In some instances, for example the welding of annealed air-hardening steels the heat-affected zone may be hard and brittle Preheating and postwelding heat-treatment may be required to avoid brittleness or restore strength To achieve higher notch impact value in steel weld metal than can be attained in a single pass, two or more passes must be deposited to take advantage of the grain refining (hence toughening) effect of each pass on the heat affected zone of the preceding pass.

Welding heat causes distortion, which can be minimized by suitable choice of fixturing, joint de-

sign, and sequence of welding. The welded part also contains internal stresses, often up to the yield strength locally in the vicinity of the weld. These may be reduced after welding by stress relief heat-treatment, which entails, in the case of steel, heating the welded structure to 1200°F and cooling slowly. See BRAZING; HEAT-TREATMENT (METALS AND ALLOYS); METALLURGY; SOLDERING.

Thermal cutting. Thermal cutting is used for a wide variety of work of the heavier type, such as structural work and shipbuilding, where the highest degree of dimensional accuracy is unimportant. Thermal cutting can be done manually or automatically and, in either case, the torch can be brought to the work. The cut can be straight-line, curved, or beveled.

The principal cutting process is oxygen cutting, also known as flame cutting, which can be used only on steels and on alloy steels low in chromium. The cutting torch tip has a central orifice for oxygen and peripheral orifices for the preheat flame (Fig. 3). The flame preheats the steel to the ignition temperature, whereupon the oxygen converts the steel to iron oxide, which issues from the bottom of the cut as a molten spray. An oxygen cutting machine can cut 1-in. steel plate at a speed of 18 in./min using an oxygen orifice of 0.0595-in. diameter and an oxygen pressure of 35 lb/in.^{2} and consuming oxygen at the rate of 150 ft^3/hr. Oxygen lancing is similar to cutting except that an oxygen pipe replaces the torch for cutting steel up to 6 ft thick.

For high-chromium alloy steels which form oxides of high melting point, powder cutting is employed. Iron powder is injected into the oxygen in one process to increase the iron oxide content of the slag and so to lower its melting point. In another process, sand, sodium bicarbonate, or calcium carbonate serves as the fluxing powder.

Two cutting processes utilize an arc. In heliarc cutting the gas stream (65% argon, 35% hydrogen) of an arc between a tungsten electrode and the work acquires high temperature and velocity in a constricting nozzle. The hot high-velocity argon

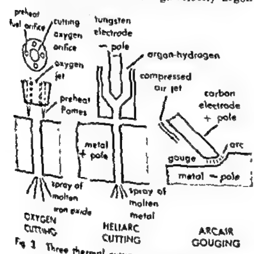


Fig. 3 Three thermal cutting processes.

produces excellent cuts in aluminum, magnesium, copper, and several other metals. For example, ½-in. aluminum plate can be cut at 150 in./min. The arcair process uses a carbon arc to melt the metal and a jet of compressed air to blow the liquid away. The process is used, for example, for gouging defects from castings. [E.E.C.]

Bibliography: American Society for Metals, *Metals Handbook*, 1948; B. E. Rossi, *Welding Engineering*, 1954.

Well

An artificial excavation made to extract water, oil, gas, brine, or other substance from the earth. Wells are the source of about one fifth of water supplies in the United States, all gas and oil, and most of the industrial brines and sulfur. Water wells are recorded in the earliest historic documents and probably originated during periods of drought when ancient man attempted to reach water at springs that had ceased to flow by digging a shallow excavation at the spring site. Later he learned to dig deeper for water where it did not issue at the land surface. The first wells were dug by hand, and some of the larger and deeper ones were provided with elaborate ramps which enabled those drawing water to walk or even drive a donkey down to the water level. On the mountain slopes around the Mediterranean sea, for thousands of years, wells (Khanats) have been constructed by tunnel-

nese appear to have developed a crude percussion drill which enabled them to dig wells a few inches in diameter, lined with bamboo, to depths of a few thousand feet. In France and Italy the drilling of wells began in the sixteenth century. In 1833 at Grenelle in Artois, France, a well was drilled to a depth of 1798 ft and lined with steel casing. Its location provided the term artesian, which originally referred only to flowing wells (see ARTESIAN SYSTEMS). Drilling for brine in the United States began in West Virginia in 1808, and for oil and gas in Pennsylvania in 1859. Drilling for water in the United States began in the 1820s but did not become common practice until the latter part of the nineteenth century. Dug wells have become almost obsolete, owing to the greater speed of drilling and greater efficiency of drilled wells. See OIL AND GAS WELLS.

Drilled wells usually are fitted with a steel tube or casing, commonly 2–36 in. in diameter, inserted in the drilled hole to the desired depth. Where the water-bearing formation is competent to stand without support, the casing is set, or finished, at the top of solid rock. Where there is danger of caving

... fine-grained sand an envelope of coarse sand or fine gravel is commonly placed around the screen. In coarser-grained aquifers the screen is

In spot welding, the sheets to be welded are held between electrodes of hard, high-conductivity copper alloy. They conduct heat away from the electrode-to-sheet contacts. The contact resistance heat at the sheet-to-sheet contact fuses the sheets together. Electrode pressure provides uniform sheet-to-sheet resistance, and prevents expulsion of metal between the sheets

Seam welding is a process for making overlapping spot welds by means of rotating electrode wheels. The joints can be made gastight and liquid-tight, as in tubing and stainless steel electric refrigerators.

In flash welding the surfaces to be welded are *held lightly in contact while a high current flows* through the few contact points. Melting occurs instantaneously at these "bridges." Violent vaporization ensues, with expulsion of small particles of hot metal as visible flash. Flashing is continued until the surfaces are coated with a layer of molten metal They are then squeezed together to form the weld.

For further information *see* FLASH WELDING; RESISTANCE WELDING; SPOT WELDING.

Gas welding A gas flame is a less-concentrated and lower-temperature source of heat than an arc. For this reason, a welding torch is often used for brazing, for welding thin material, and for applications requiring low gradients of temperature to avoid cracking, as in welding cast iron. The oxyacetylene flame, using a mixture of equal volumes of oxygen and acetylene (C_2H_2), has a higher temperature than any other commercially available fuel gas combination. High flame temperature is essential for the rapid localized heating required in welding Several sizes and types of torch tips are available to secure flames of different intensities and dispersions. The ratio of oxygen to acetylene governs the chemical action of the flame. If the ratio is higher than unity, for example 1.1:1, the flame is oxidizing and is useful for preventing hydrogen porosity in brass. With a ratio below unity, 0 95:1, an acetylene feather appears in the flame, carbon is picked up by molten steel, and mill scale on the surface of steel is reduced A neutral flame with a ratio of 1:1 is used for steel Flux is not used in welding steel, but is necessary for gas welding nonferrous metals. The use of other fuel gases, such as hydrogen, is restricted to the welding of metals of low melting point, such as lead.

Other fusion-welding processes Among other fusion-welding processes are the induction, thermit, and hot air processes. The heat for induction welding is obtained from the resistance of the work to induced high-frequency current. *Proximity effect* concentrates the current in the edges to be joined, as in the manufacture of metal radio tubes In ther- [illegible] ylene, are welded by means of hot air torches. the current of hot air melting the edges that are to be joined

Nonfusion processes. In these processes, the surfaces to be welded coalesce in the solid state by application of pressure with or without heat. It is essential that the surfaces be free from films, such as oxides. Resistance upset welding uses contact resistance heat to bring the surface to a high temperature but below the melting point. Unlike flash welding, the surfaces are squeezed tightly together throughout the process, which upsets, or changes the dimensions of, the metal close to the weld In forge welding, the parts are heated in a furnace and pressure is applied usually by hammer blows. The heat for gas-pressure and induction-pressure welding is derived from flames and induced electric currents, respectively. *Cold welding* is performed by pressure alone without heat, and is applicable particularly to soft metals, such as aluminum. Ultrasonic welding also requires no heat and only light pressure; the atomic movement required for coalescence is stimulated by ultrasonic vibrations.

Welding metallurgy. The effects of welding on a product often govern its suitability for service. The weld metal, the zone close to the weld (called the heat-affected zone), and the thermal stresses and strains caused by welding may combine to yield a product different in strength, ductility, leaktightness, and dimensional stability than the unwelded base metal. Weld metal frequently is compared with cast metal with which it shares the characteristic of freezing in columnar dendritic pattern in the direction opposite to the direction of withdrawal of heat Unlike cast metal, weld metal usually has a free surface exposed to the air, and it is free from shrinkage cavities. Weld metal freezes rapidly compared with most castings and is comparatively fine-grained and free from chemical segregation Porosity must be guarded against. It is a result of evolution of gas, such as nitrogen from steel, during freezing of the gas-saturated weld pool. Porosity also may result from a chemical reaction in the weld metal, such as the reaction between carbon and oxygen in liquid steel. Hot cracking of the weld metal during the last stages of solidification may be prevented either by fixturing so as to avoid strain during freezing, or by adding suitable alloying elements to the weld, as in the addition of manganese to steel weld metal to combine with sulfur

The heat-affected zone may be softer than the base metal, especially in alloys that have been cold-rolled or heat-treated for high strength In some instances, for example the welding of annealed air-hardening steels. the heat-affected zone may be hard and brittle. Preheating and postwelding heat-treatment may be required to avoid brittleness or restore strength. To achieve higher notch impact value in steel weld metal than can be attained in a single pass, two or more passes must be deposited to take advantage of the grain refining (hence toughening) effect of each pass on the heat-affected zone of the preceding pass.

Welding heat causes distortion, which can be minimized by suitable choice of fixturing, joint de-

mine deviation as well as drift by photographing plumb bob position against a compass background. A gyroscopic compass may be used in holes of large enough diameter An instrument for use in diamond drill holes even as small as 1½ in uses a time delay device to lock a compass needle and also lock the plumb bob type compass case to graduated rings, which are oriented with the axis of the hole

Electric well logging. All logging methods utilizing electrical recording instruments run in a drill hole are frequently termed electric well logging More specifically the term describes logging methods recording variations in electrical potential due to the natural electrical properties of subsurface formations . . .

electr

recip

. . . porosity, and fluid content.

Basic electric logging produces a spontaneous potential curve plus one or a set of resistivity curves, in uncased holes containing an electrically conducting fluid

Spontaneous potential logging. This measures the . . .

Resistivity logging Variations in potential under the influence of an electric current introduced through a current electrode run in the drill hole are measured and recorded by this method Three basic systems of resistivity logging are used as illustrated

The number and placing of recording electrodes is optional, permitting a wide range of readings obtained simultaneously on one trip of the instrument down the hole Multi conductor cables are used to obtain the different readings, or a single conductor may be used with a frequency modulation system The monoelectrode system is most frequently used for correlation work in core drill holes; normal and lateral electrode systems, in oil and gas well logging.

If the well contains no fluid or a nonconducting fluid, "scratcher" type, wall-contacting electrodes may be tried

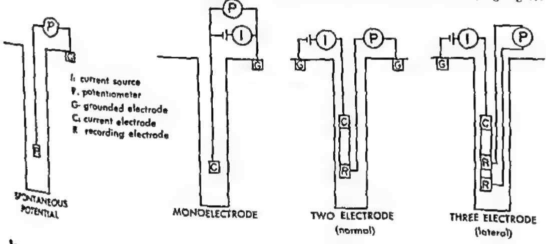
Resistivity micro-logging. Three contact electrodes are placed close together in a pad that is held against the wall of the hole by springs to obtain localized resistivity readings. Such a technique is used to distinguish between permeable and nonpermeable formations, on the principle that permeable formations should give lower resistivity readings due to both presence of mud cake and invasion of fluid. It is one of the special logging techniques for recognition of impermeable limestones

Focused-current logging. Auxiliary power electrodes are used in this method to create interference and focus the current into a disk shape perpendicular to the drill hole. The focused current increases the power of resolution and overcomes damping effects of the drilling fluid Both resistivity and conductivity curves may be obtained, the conductivity curve being useful in zones of low resistivity because it amplifies small changes

Electromagnetic (induction) logging Only holes above fluid level, as in cable tool holes or where the type of mud used has very low conductivity, can be investigated by this method A transmitter coil sets up an alternating electromagnetic field, which induces eddy currents within the formations. The eddy currents establish a secondary electromagnetic field which in turn induces an electromotive force in a receiver coil placed in the drill hole above and spaced from the transmitter coil. The electromotive force recorded by the receiver coil is a measure of the conductivity of the formations, which is inversely proportional to the . . .

Radio. . . natural or . . . nations It has the advantages of being operable in cased or uncased holes, whether filled with fluid or not

Gamma-ray logging Rock character is reflected by variations in the intensity of radioactive gamma rays emitted from all formations in varying degrees



Basic patterns of electric well logging

directly in contact with the aquifer. In either case the construction includes a considerable period of pumping, surging, or other treatment (called well development), during which the finer particles of the formation are drawn into the well and removed. This process commonly increases the initial yield of the well substantially, in some cases by several times. Most wells of large capacity are equipped with pumps of the deep-well turbine type to lift the water to the surface. However, where the water surface is less than 20 ft below the land surface they may be equipped with horizontal centrifugal pumps or other pumps of the suction type. In general, the pump is selected to fit the needs of the individual well. Many small-capacity farm and ranch wells are fitted with reciprocating pumps powered by windmills. Domestic wells are commonly equipped with electrically powered reciprocating or jet pumps. A few industrial wells, mostly old, are pumped by air lift. *See* PUMPING MACHINERY.

When a well is pumped, the pressure head at the well is lowered and a hydraulic gradient toward the well is established which causes water to flow toward the well. This lowering of head is called drawdown (*see* GROUND WATER). One commonly used method for judging the ultimate yield of a well is to measure the yield per unit of drawdown. *This is the specific capacity of the well*, commonly expressed in gallons per minute per foot of drawdown. [A.N.S.]

Well logging (mineral)

The technique of analyzing and recording the character of the formations penetrated by a drill hole in mineral exploration and exploitation work. *See* BORING AND DRILLING, MINERAL.

Drill-hole logs (records) are compilations of data obtained from (1) recordings of the action of drill and auxiliary equipment, (2) examination of samples produced by the drilling, in the form of cores, cuttings, and sludge, and (3) examination of the wall of the hole, by visual and geophysical methods.

Mineral drilling almost always produces samples in some form as a basis for logging. Core drilling is the most accurate because it recovers unaltered samples in the form of cores. In all other drilling methods, cuttings of some sort are produced as the hole progresses. But even where 100% core recovery is obtained, other methods are needed for complete logging. For example, the position of natural groundwater level cannot be determined from the core. Also, geophysical logging methods may disclose mineralization close to but not in the direct path of the drill bit.

Driller's log. This is a record of all occurrences during drilling that might help in a complete logging of the hole or in determining the cost of the drilling. Observations of changes in rate of drilling, pump pressure, action of the tools, wear on the tools, and color of the circulating fluid may furnish the best clues as to depths at which formation changes occurred. On core drills the driller is usually required to log the formations whether or not the cores are to be analyzed by a geologist or laboratory technician, as the driller can establish the depths of stratigraphic changes where core recovery is incomplete. On large drills it is common practice to use instruments to assist the driller, such as rate of penetration recorders, and sometimes pump pressure, flow rate, and tool weight recorders.

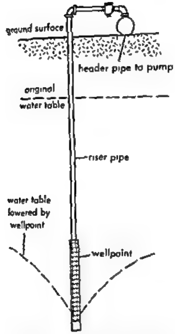
Core logging. Investigation and recording of this kind may be as complete as the project warrants, from visual recognition of valueless rocks to the multitude of physical and chemical tests to which a mineral specimen may be subjected. Visual examination is the most important phase of core logging, as any analysis must be based on recognition of rock type, identification of mineral content, and determination of physical condition. Visual examination may include microscopic study of grains and fossils, and thin sections under polarized light. It is customary practice to split a diamond drill core, sending one-half or one-fourth to a laboratory and keeping at least one-half for permanent record and possible additional study. Core orientation is possible by laboratory determination of magnetic polarity. Samples or so-called cores may be obtained from the wall of the hole in soft formations by side-wall coring methods.

Where core recovery may be low it is customary to collect the sludge formed by both cuttings and ground up core. Various tables have been developed for combining core and sludge analyses for interpretation.

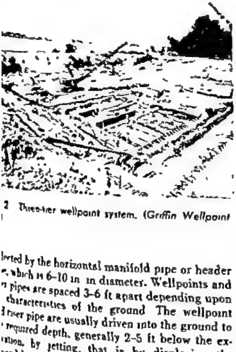
Cuttings-analysis logging. An interpretation of the driller's log, commonly supplemented by other logging methods, is the basis of this technique. Cuttings can only be ascribed to approximate depths, as they are either in a mixture from a section of the hole, as in cable tool drilling, or there is a time delay in their reaching the surface. They are also subject to contamination by caving of the walls of the hole at a higher level. When cuttings are recovered from drilling fluid only the coarser particles can be settled out or screened out. Despite the difficulties, cuttings analysis furnishes much valuable data and combined with other logging methods may furnish sufficient data for certain projects at less cost than core drilling.

Visual wall logging. *This may be done by actual* examination of the wall of the hole by a geologist in drilled shafts and in inspection holes drilled 30 in. or larger in diameter. Logging of the wall has certain advantages over core logging in that sedimentation and fracture patterns may be recognized with true orientation. In smaller holes visual wall examination may be made with a moving picture bore hole camera or with a closed-circuit television camera. *See* CLOSED-CIRCUIT TELEVISION.

Directional surveying. The drift of a hole from the vertical and its deviation in azimuth are logged during this procedure. Drift recorders utilize the fluid level principle, such as the use of hydrofluoric acid to etch a horizontal line on a glass tube, or the plumb bob principle. Photoclinometers deter-



1 Components of a wellpoint system


2 Three-tier wellpoint system. (Griffin Wellpoint

...ected by the horizontal manifold pipe or header ..., which is 6–10 in in diameter. Wellpoints and ...r pipes are spaced 3–6 ft apart depending upon ... characteristics of the ground The wellpoint ...d riser pipe are usually driven into the ground to ... required depth, generally 2–5 ft below the ex...vation, by jetting, that is by displacing the ...ound below it with water under pressure. Where the ground-water stratum is overlaid with ...d clay or heavy gravel, jetting may not be able ... pierce the tough material and a special boring ...ipment is used to carry the wellpoint down The pumps are located above the water table and ...vert the water from the header pipes for dis...charge away from the excavation area The pumps ... centrifugal pumps, with vacuum pumps to in...crease the suction lift. The suction lift in each ...point rarely exceeds 18–22 ft Consequently, ...here the depth of the wellpoint below ground level exceeds this depth, it is necessary to install two or more tiers of wellpoint systems, each with its own assembly of wellpoints, header pipes, and pumps. A three-tier wellpoint system is shown in Fig. 2 [W.H.]

West Indies

An archipelago, including the Bahamas, the Greater Antilles, the Lesser Antilles, and other islands, curving 2500 mi from Yucatan Peninsula and southeastern Florida to northern Venezuela and enclosing the Caribbean Sea. Situated between latitude 10° and 27° N and longitude 59° and 85° W, in the zone of the northeast trade winds, the West Indies have a subtropical and predominantly oceanic climate, with even warmth and steady breezes. Temperatures vary little from season to season, ranging from means of 80–85°F in July to 70–78°F in January at sea level Freezing is unknown, and the hottest temperatures rarely exceed 90°F. Precipitation ranges from a low of 25–50 in. a year on low-lying islands and drier coasts up to 300 in. on the highest peaks, which are almost perpetually cloud-capped. At lower elevations, rainfall is erratic from year to year and from season to season, but reaches a maximum in the summer and fall, when the northeast trades are replaced by light, variable winds. This is also the season of hurricanes, destructive tropical cyclones which sweep west and northwest across the Caribbean, sparing only the southernmost islands. The winter months are generally dry, and there is frequently a shorter dry season in July or August.

The flora and fauna of the islands have been largely transformed by human occupancy; today sugar cane and cotton occupy much of the lowlands, while bananas, tobacco, cacao, citrus, coffee, and subsistence crops mantle the foothills and mountain slopes Mangroves and other salt-water vegetation still fringe most of the islands; cactus, thorn bush, and dry scrub occur on protected dry leeward sites, while other low-lying areas are savannas with scattered palms and pines. Higher up, where droughts are briefer and less intense, mixed conifer and broadleaf forests are to be found, with rain forest above. Windswept and sodden summits support palm brake and elfin woodland.

The West Indian flora is chiefly derived from Central and South America, but there are a number of endemic species, notably palms; many mainland plants failed to colonize the islands, which are floristically poor. The effect of isolation is especially evident in the meager character of West Indian fauna Animal species as a whole are limited; there are few mammals and no large ones.

Population figures are 1957 estimates unless otherwise noted.

Bahama Islands. The Bahamas (4375 sq mi, population 131,000), together with the Turks and Caicos Islands (166 sq mi, population 7000) to the southwest, make up about 3000 islands, rocks, and cays stretching more than 600 mi southeastward from Grand Bahama Island, 50 mi east of

because of the presence of uranium, thorium, or potassium minerals, even though very minute in amount. Gamma rays are produced when radioactive substances emit alpha and beta particles. *Gamma rays have certain properties which make* possible the recording of their intensity. The Geiger-Müller counter utilizes their property of causing ionization and momentary conductivity in a normally nonconducting gas in a tube between two electrodes under high potential. The scintillometer utilizes their property of causing certain crystals to fluoresce.

Gamma-ray curves are helpful in stratigraphic correlation, shales generally exhibiting a higher degree of radioactivity than sandstones or limestones. Such curves are invaluable in prospecting for radioactive minerals, which may be close to, but *missed by, the drill hole.*

Neutron logging. The concentration of hydrogen atoms in the rock formations is measured and thereby fluid content and porosity is indicated. Neutron logging is of two types, depending on the type of detector employed. Both use a source of fast neutrons such as a mixture of beryllium and a radioactive element. Neutron-gamma-ray logging utilizes a gamma-ray detector to measure increase in gamma-ray emission caused by capture of fast neutrons by hydrogen nuclei. Neutron-neutron logging utilizes a slow neutron detector. Fast neutrons produced by one part of the instrument are deflected and slowed by collision with hydrogen atoms in the formation and some return as slow neutrons to the recording part of the instrument.

Other significant techniques. Eight other types are of considerable use.

Caliper logging. Variations in hole diameter due to erosion of the formations or to mud cake deposition are measured by this method. Spring-loaded wall contact pads or fingers measure the diameter as the instrument is dragged up the wall. A rheostat is usually connected to the arms to record their expansion and contraction by current fluctuations in the recording circuit.

Dip logging. This determines the angle and azimuth of dip of nonhorizontal formations. One instrument used is a calipering device with three arms spaced at 120° to record, at three points on the wall, the variation in depth of persistent erosional features of a bed cut at an angle by the drill hole. A second type of dipmeter uses the three arms to measure the three depths of change in resistivity at bedding contacts. Another type operates on the principle that resistivity is lower in the direction of the bedding planes. This resistivity can be determined by rotating two potential recording electrodes in a plane perpendicular to the axis of the drill hole. Azimuth may be determined by orientation *in shallow holes* or by photoclinometer or induction compass.

Sonic logging. A pulse-echo system is utilized to measure the distance between the instrument and a sound-reflecting surface. The method is not operable at distances of less than 1½ ft, but is useful in measuring the size of cavities around brine wells and the capacities of underground LPG (liquidified petroleum gas) storage chambers.

Acoustic logging. Variations in sound wave transmission are measured between a high energy source and a receiver spaced vertically in a drill hole.

Geophone (velocity) logging. A seismic method, using a geophone (seismometer) in the drill hole, indicates formation changes by comparison of seismic readings at different depths with readings made by a reference seisomometer on the surface.

Geothermal (temperature) logging. Measured variations of temperature with depth are useful in interpreting resistivity curves since the resistivity of rocks varies with their temperatures and there is a natural progression to higher temperatures with depth. It may also be used to locate points of en*try or loss of water, entry of gases, and to deter*mine depths of cementation behind casing.

Photoelectric logging. This is a technique used to determine points of water entry into a drill hole. A light-absorbing dye is distributed throughout the water in a drill hole. The fluid level is then lowered to induce water entry and a photoelectric tool run to measure light intensities and determine where dilution has occurred.

Geochemical well logging. Characteristics of cores and samples are observed and analyzed as in geochemical surface prospecting so that trace element patterns may indicate possible greater concentrations at depth and the advisability of drilling deeper. Spectroscopic methods as well as delicate chemical methods may be used for geochemical analysis. *See* GEOPHYSICAL EXPLORATION; PROSPECTING. [F.C.ST.]

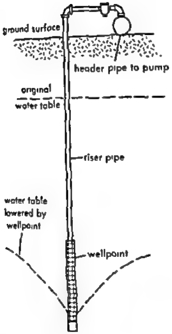
Wellpoint systems

A method of keeping an excavated area dry by intercepting the flow of ground water with pipe wells located around the excavation area. Intercepting the flow before it reaches the excavated area also improves the stability of the edge of the excavation, permitting steeper bank slopes and often eliminating the need for supporting or shoring the banks. *See* CONSTRUCTION METHODS.

Wellpoint systems are most effective in coarse-grained soils, such as gravel or sand. They are not effective in fine soils, such as silts and clays, where the small size of the pores between grains restricts the flow of water. For a discussion of dewatering in fine-grained soils, *see* ELECTROOSMOTIC DEWATERING.

The basic components of a wellpoint system are the wellpoint, the riser pipe, the header pipe or manifold, and the pump.

The wellpoint consists of a perforated pipe about 4 ft long and about 2 in. in diameter. It is equipped with a ball valve to regulate the flow of water, a screen to prevent the entry of sand during pumping, and a jetting tip. The steel riser pipe to which the wellpoint is attached is slightly smaller in diameter than the wellpoint. Its purpose is to bring the ground water to the surface, where it is



1 Components of a wellpoint system

2 Three-tier wellpoint system (Griffin Wellpoint
)

llected by the horizontal manifold pipe or header pe, which is 6–10 in. in diameter. Wellpoints and ser pipes are spaced 3–6 ft apart depending upon e characteristics of the ground The wellpoint nd riser pipe are usually driven into the ground to e required depth, generally 2–5 ft below the exavation, by jetting, that is by displacing the round below it with water under pressure

Where the ground water stratum is overlaid with ard clay or heavy gravel, jetting may not be able o pierce the tough material and a special boring ttachment is used to carry the wellpoint down

The pumps are located above the water table and ollect the water from the header pipes for discharge away from the excavation area The pumps re centrifugal pumps, with vacuum pumps to increase the suction lift. The suction lift in each wellpoint rarely exceeds 18–22 ft. Consequently, where the depth of the wellpoint below ground level exceeds this depth, it is necessary to install two or more tiers of wellpoint systems, each with its own assembly of wellpoints, header pipes, and pumps. A three-tier wellpoint system is shown in Fig. 2. [W.H.]

West Indies

An archipelago, including the Bahamas, the Greater Antilles, the Lesser Antilles, and other islands, curving 2500 mi from Yucatan Peninsula and southeastern Florida to northern Venezuela and enclosing the Caribbean Sea. Situated between latitude 10° and 27° N and longitude 59° and 85° W, in the zone of the northeast trade winds, the West Indies have a subtropical and predominantly oceanic climate, with even warmth and steady breezes Temperatures vary little from season to season, ranging from means of 80–85°F in July to 70–78°F in January at sea level Freezing is unknown, and the hottest temperatures rarely exceed 90°F. Precipitation ranges from a low of 25–50 in a year on low-lying islands and drier coasts up to 300 in. on the highest peaks, which are almost perpetually cloud-capped At lower elevations, rainfall is erratic from year to year and from season to season, but reaches a maximum in the summer and fall, when the northeast trades are replaced by light, variable winds. This is also the season of hurricanes, destructive tropical cyclones which sweep west and northwest across the Caribbean, sparing only the southernmost islands. The winter months are generally dry, and there is frequently a shorter dry season in July or August.

The flora and fauna of the islands have been largely transformed by human occupancy; today sugar cane and cotton occupy much of the lowlands, while bananas, tobacco, cacao, citrus, coffee, and subsistence crops mantle the foothills and mountain slopes Mangroves and other salt-water vegetation still fringe most of the islands; cactus, thorn bush, and dry scrub occur on protected dry leeward sites, while other low-lying areas are savannas with scattered palms and pines. Higher up, where droughts are briefer and less intense, mixed conifer and broadleaf forests are to be found, with rain forest above. Windswept and sodden summits support palm brake and elfin woodland.

The West Indian flora is chiefly derived from Central and South America, but there are a number of endemic species, notably palms; many mainland plants failed to colonize the islands, which are floristically poor The effect of isolation is especially evident in the meager character of West Indian fauna. Animal species as a whole are limited, there are few mammals and no large ones

Population figures are 1957 estimates unless otherwise noted

Bahama Islands. The Bahamas (4375 sq mi, population 131,000), together with the Turks and Caicos Islands (166 sq mi, population 7000) to the southwest, make up about 3000 islands, rocks, and cays stretching more than 600 mi southeastward from Grand Bahama Island, 50 mi east of

because of the presence of uranium, thorium, or potassium minerals, even though very minute in amount. Gamma rays are produced when radioactive substances emit alpha and beta particles. Gamma rays have certain properties which make possible the recording of their intensity. The Geiger-Müller counter utilizes their property of causing ionization and momentary conductivity in a normally nonconducting gas in a tube between two electrodes under high potential. The scintillometer utilizes their property of causing certain crystals to fluoresce.

Gamma-ray curves are helpful in stratigraphic correlation, shales generally exhibiting a higher degree of radioactivity than sandstones or limestones. Such curves are invaluable in prospecting for radioactive minerals, which may be close to, but missed by, the drill hole.

Neutron logging. The concentration of hydrogen atoms in the rock formations is measured and thereby fluid content and porosity is indicated. Neutron logging is of two types, depending on the type of detector employed. Both use a source of fast neutrons such as a mixture of beryllium and a radioactive element. Neutron-gamma-ray logging utilizes a gamma-ray detector to measure increase in gamma-ray emission caused by capture of fast neutrons by hydrogen nuclei. Neutron-neutron logging utilizes a slow neutron detector. Fast neutrons produced by one part of the instrument are deflected and slowed by collision with hydrogen atoms in the formation and some return as slow neutrons to the recording part of the instrument.

Other significant techniques. Eight other types are of considerable use.

Caliper logging. Variations in hole diameter due to erosion of the formations or to mud cake deposition are measured by this method. Spring-loaded wall contact pads or fingers measure the diameter as the instrument is dragged up the wall. A rheostat is usually connected to the arms to record their expansion and contraction by current fluctuations in the recording circuit.

Dip logging. This determines the angle and azimuth of dip of nonhorizontal formations. One instrument used is a calipering device with three arms spaced at 120° to record, at three points on the wall, the variation in depth of persistent erosional features of a bed cut at an angle by the drill hole. A second type of dipmeter uses the three arms to measure the three depths of change in resistivity at bedding contacts. Another type operates on the principle that resistivity is lower in the direction of the bedding planes. This resistivity can be determined by rotating two potential recording electrodes in a plane perpendicular to the axis of the drill hole. Azimuth may be determined by orientation in shallow holes or by photoclinometer or induction compass.

Sonic logging. A pulse-echo system is utilized to measure the distance between the instrument and a sound-reflecting surface. The method is not operable at distances of less than $1\frac{1}{2}$ ft, but is useful in measuring the size of cavities around brine wells and the capacities of underground LPG (liquidified petroleum gas) storage chambers.

Acoustic logging. Variations in sound wave transmission are measured between a high energy source and a receiver spaced vertically in a drill hole.

Geophone (velocity) logging. A seismic method, using a geophone (seismometer) in the drill hole, indicates formation changes by comparison of seismic readings at different depths with readings made by a reference seisomometer on the surface.

Geothermal (temperature) logging. Measured variations of temperature with depth are useful in interpreting resistivity curves since the resistivity of rocks varies with their temperatures and there is a natural progression to higher temperatures with depth. It may also be used to locate points of entry or loss of water, entry of gases, and to determine depths of cementation behind casing.

Photoelectric logging. This is a technique used to determine points of water entry into a drill hole. A light-absorbing dye is distributed throughout the water in a drill hole. The fluid level is then lowered to induce water entry and a photoelectric tool run to measure light intensities and determine where dilution has occurred.

Geochemical well logging. Characteristics of cores and samples are observed and analyzed as in geochemical surface prospecting so that trace element patterns may indicate possible greater concentrations at depth and the advisability of drilling deeper. Spectroscopic methods as well as delicate chemical methods may be used for geochemical analysis. *See* GEOPHYSICAL EXPLORATION; PROSPECTING. [F.C.ST.]

Wellpoint systems

A method of keeping an excavated area dry by intercepting the flow of ground water with pipe wells located around the excavation area. Intercepting the flow before it reaches the excavated area also improves the stability of the edge of the excavation, permitting steeper bank slopes and often eliminating the need for supporting or shoring the banks. *See* CONSTRUCTION METHODS.

Wellpoint systems are most effective in coarse-grained soils, such as gravel or sand. They are not effective in fine soils, such as silts and clays, where the small size of the pores between grains restricts the flow of water. For a discussion of dewatering in fine-grained soils, *see* ELECTROOSMOTIC DEWATERING.

The basic components of a wellpoint system are the wellpoint, the riser pipe, the header pipe or manifold, and the pump.

The wellpoint consists of a perforated pipe about 4 ft long and about 2 in. in diameter. It is equipped with a ball valve to regulate the flow of water, a screen to prevent the entry of sand during pumping, and a jetting tip. The steel riser pipe to which the wellpoint is attached is slightly smaller in diameter than the wellpoint. Its purpose is to bring the ground water to the surface, where it is

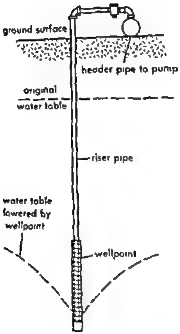


Fig 1 Components of a wellpoint system

Fig. 2 Three tier wellpoint system (Griffin Wellpoint Co)

... of the ground. The wellpoint and riser pipe are usually driven into the ground to the required depth, generally 2–5 ft below the excavation, by *jetting*, that is by displacing the ground below it with water under pressure.

Where the ground water stratum is overlaid with hard clay or heavy gravel, jetting may not be able to pierce the tough material and a special boring attachment is used to carry the wellpoint down.

The pumps are located above the water table and collect the water from the header pipes for discharge away from the excavation area. The pumps are centrifugal pumps, with vacuum pumps to increase the suction lift. The suction lift in each wellpoint rarely exceeds 18–22 ft. Consequently, where the depth of the wellpoint below ground level exceeds this depth, it is necessary to install two or more tiers of wellpoint systems, each with its own assembly of wellpoints, header pipes, and pumps. A three-tier wellpoint system is shown in Fig. 2. [W.H.]

West Indies

An archipelago, including the Bahamas, the Greater Antilles, the Lesser Antilles, and other islands, curving 2500 mi from Yucatan Peninsula and southeastern Florida to northern Venezuela and enclosing the Caribbean Sea. Situated between latitude 10° and 27° N and longitude 59° and 85° W, in the zone of the northeast trade winds, the West Indies have a subtropical and predominantly oceanic climate, with even warmth and steady breezes. Temperatures vary little from season to season, ranging from means of 80–85°F in July to 70–78°F in January at sea level. Freezing is unknown, and the hottest temperatures rarely exceed 90°F. Precipitation ranges from a low of 25–50 in. a year on low-lying islands and drier coasts up to 300 in. on the highest peaks, which are almost perpetually cloud-capped. At lower elevations, rainfall is erratic from year to year and from season to season, but reaches a maximum in the summer and fall, when the northeast trades are replaced by light, variable winds. This is also the season of hurricanes, destructive tropical cyclones which sweep west and northwest across the Caribbean, sparing only the southernmost islands. The winter months are generally dry, and there is frequently a shorter dry season in July or August.

The flora and fauna of the islands have been largely transformed by human occupancy; today sugar cane and cotton occupy much of the lowlands, while bananas, tobacco, cacao, citrus, coffee, and subsistence crops mantle the foothills and mountain slopes. Mangroves and other salt-water vegetation still fringe most of the islands; cactus, thorn bush, and dry scrub occur on protected dry leeward sites, while other low-lying areas are sa-

... with rain forest above. Windswept and sodden summits support palm brake and elfin woodland.

The West Indian flora is chiefly derived from Central and South America, but there are a number of endemic species, notably palms; many mainland plants failed to colonize the islands, which are floristically poor. The effect of isolation is especially evident in the meager character of West Indian fauna. Animal species as a whole are limited; there are few mammals and no large ones.

Population figures are 1957 estimates unless otherwise noted.

Bahama Islands. The Bahamas (4375 sq mi; population 131,000), together with the Turks and Caicos Islands (166 sq mi; population 7000) to the southwest, make up about 3000 islands, rocks, and cays stretching more than 600 mi southeastward from Grand Bahama Island, 50 mi east of

because of the presence of uranium, thorium, or potassium minerals, even though very minute in amount. Gamma rays are produced when radioactive substances emit alpha and beta particles. Gamma rays have certain properties which make possible the recording of their intensity. The Geiger-Muller counter utilizes their property of causing ionization and momentary conductivity in a normally nonconducting gas in a tube between two electrodes under high potential. The scintillometer utilizes their property of causing certain crystals to fluoresce.

Gamma-ray curves are helpful in stratigraphic correlation, shales generally exhibiting a higher degree of radioactivity than sandstones or limestones. Such curves are invaluable in prospecting for radioactive minerals, which may be close to, but missed by, the drill hole.

Neutron logging. The concentration of hydrogen atoms in the rock formations is measured and thereby fluid content and porosity is indicated. Neutron logging is of two types, depending on the type of detector employed. Both use a source of fast neutrons such as a mixture of beryllium and a radioactive element. Neutron-gamma-ray logging utilizes a gamma-ray detector to measure increase in gamma-ray emission caused by capture of fast neutrons by hydrogen nuclei. Neutron-neutron logging utilizes a slow neutron detector. Fast neutrons produced by one part of the instrument are deflected and slowed by collision with hydrogen atoms in the formation and some return as slow neutrons to the recording part of the instrument.

Other significant techniques. Eight other types are of considerable use.

Caliper logging. Variations in hole diameter due to erosion of the formations or to mud cake deposition are measured by this method. Spring-loaded wall contact pads or fingers measure the diameter as the instrument is dragged up the wall. A rheostat is usually connected to the arms to record their expansion and contraction by current fluctuations in the recording circuit.

Dip logging. This determines the angle and azimuth of dip of nonhorizontal formations. One instrument used is a calipering device with three arms spaced at 120° to record, at three points on the wall, the variation in depth of persistent erosional features of a bed cut at an angle by the drill hole. A second type of dipmeter uses the three arms to measure the three depths of change in resistivity at bedding contacts. Another type operates on the principle that resistivity is lower in the direction of the bedding planes. This resistivity can be determined by rotating two potential recording electrodes in a plane perpendicular to the axis of the drill hole. Azimuth may be determined by orientation in shallow holes or by photoclinometer or induction compass.

Sonic logging. A pulse-echo system is utilized to measure the distance between the instrument and a sound-reflecting surface. The method is not operable at distances of less than $1\frac{1}{2}$ ft, but is useful in measuring the size of cavities around brine wells and the capacities of underground LPG (liquidified petroleum gas) storage chambers.

Acoustic logging. Variations in sound wave transmission are measured between a high energy source and a receiver spaced vertically in a drill hole.

Geophone (velocity) logging. A seismic method, using a geophone (seismometer) in the drill hole indicates formation changes by comparison of seismic readings at different depths with readings made by a reference seisomometer on the surface.

Geothermal (temperature) logging. Measured variations of temperature with depth are useful in interpreting resistivity curves since the resistivity of rocks varies with their temperatures and there is a natural progression to higher temperatures with depth. It may also be used to locate points of entry or loss of water, entry of gases, and to determine depths of cementation behind casing.

Photoelectric logging. This is a technique used to determine points of water entry into a drill hole. A light-absorbing dye is distributed throughout the water in a drill hole. The fluid level is then lowered to induce water entry and a photoelectric tool run to measure light intensities and determine where dilution has occurred.

Geochemical well logging. Characteristics of cores and samples are observed and analyzed as in geochemical surface prospecting so that trace element patterns may indicate possible greater concentrations at depth and the advisability of drilling deeper. Spectroscopic methods as well as delicate chemical methods may be used for geochemical analysis. *See* GEOPHYSICAL EXPLORATION; PROSPECTING. [F.C.ST.]

Wellpoint systems

A method of keeping an excavated area dry by intercepting the flow of ground water with pipe wells located around the excavation area. Intercepting the flow before it reaches the excavated area also improves the stability of the edge of the excavation, permitting steeper bank slopes and often eliminating the need for supporting or shoring the banks. *See* CONSTRUCTION METHODS.

Wellpoint systems are most effective in coarse-grained soils, such as gravel or sand. They are not effective in fine soils, such as silts and clays, where the small size of the pores between grains restricts the flow of water. For a discussion of dewatering in fine-grained soils, *see* ELECTROOSMOTIC DEWATERING.

The basic components of a wellpoint system are the wellpoint, the riser pipe, the header pipe or manifold, and the pump.

The wellpoint consists of a perforated pipe about 4 ft long and about 2 in. in diameter. It is equipped with a ball valve to regulate the flow of water, a screen to prevent the entry of sand during pumping, and a jetting tip. The steel riser pipe to which the wellpoint is attached is slightly smaller in diameter than the wellpoint. Its purpose is to bring the ground water to the surface, where it is

scrub steppe of pine and palmetto interspersed with short grass. Among numerous offshore islands, the Isle of Pines (1182 sq mi), to the southwest, is largest.

Jamaica. This British territory (4411 sq mi; population 1,611,000) is built up on an east-west trending igneous and metamorphic core, which reaches its greatest elevation in the Blue Mountains (Blue Mountain Peak, 7402 ft) at the eastern end of the island and is exposed at lower heights in the center. These steep and rugged mountains are subject to landslides and severe erosion where deforested. To the west and on their north and south flanks they are overlain, at about 3000 ft above sea level, by a limestone plateau. But the limestone is so intricately dissected that little remains of the plateau surface. Karst topography is widespread, and in the Cockpit Country of north-central Jamaica, takes the form of sinkholes 500 ft deep and a quarter of a mile across. A by-product of weathering in some of the limestone glades is bauxite ore, of which Jamaica is estimated to possess 600,000,000 commercially utilizable tons. Thanks to their high alumina content, their surface occurrence, and their proximity to North American reduction plants, the Jamaican deposits are currently the world's most productive (4,634,000 tons in 1957). Level land is practically confined to alluvial plains, of which the largest is along the south central coast. Jamaica belongs to The West Indies federation.

Hispaniola. With a total area of 30,000 sq mi, the eastern two-thirds of this island is the Dominican Republic (population 2,698,000), the western third is Haiti (population 3,384,000). It is the most mountainous and structurally complex of the Greater Antilles. The Cordillera Central traverses the island from southeast to northwest at 8000–9000 ft, culminating in Pico Trujillo (about 10,200 ft), the highest elevation in the West Indies. There are two other major ranges. In the north, entirely within the Dominican Republic, the 4000-ft Cordillera Septentrionale runs eastward into the Peninsula of Samaná. In the southwest, another range forms the Tiburón Peninsula of Haiti and extends eastward into the Sierra Baharuco of the Dominican Republic; though narrower, it is almost as high as the Cordillera Central, reaching 7920 ft in the western Massif de la Hotte and 8793 ft toward the east in the Massif de la Selle. Between the southwest and central ranges runs a deep structural trench, the Cul de Sac, a continuation of the Cayman Trough between Cuba and Jamaica. The Cul de Sac contains two brackish lakes, of which the easternmost—Lake Enriquillo in the Dominican Republic—is 160 ft below sea level.

Besides the Cul de Sac, the only extensive level areas in Hispaniola are the plain separating the northern and central ranges (Plaine du Nord in Haiti, the Cibão in the Dominican Republic), the extensive Seibo lowland in southeastern Dominican Republic, and the plain of the Artibonite River at the western end of the Cordillera Central. Above the latter at about 1000 ft lies a rolling intermont basin, the Plaine Centrale. The largest offshore island, Ile de la Gonave (287 sq mi), lies in the Golfe des Gonaïves, almost encircled by the two western peninsulas.

Hispaniola's climate is less oceanic than that of the other West Indies. The trade winds fail to penetrate to many sheltered locations, which warm up like continental areas in summer. Because of the heat, some of the plains are subject to drought even though they receive considerable rainfall; the Cul de Sac gets almost 50 in. annually, but evaporation is so intense that the vegetation is of semiarid type.

Puerto Rico. The top of an eroded mountain core forms the structure of this island (3423 sq mi; population 2,281,000). The highest part rises to 4400 ft in Cerro de Punta. To the south is a narrow, arid coastal plain. The northern coastal plain, broader and comparatively well-watered, is separated from the mountains by dissected limestone terraces, except in the northeast, where the Sierra de Luquillo (3500 ft), an outlier of the central range, reaches close to the sea. Little of the natural vegetation remains anywhere. A Spanish dependency until 1898, Puerto Rico together with its offshore islands, Mona (west), Vieques, and Culebra (east), has since been an American possession, and is now a self-governing commonwealth under the United States.

Virgin Islands. These small islands with a total area of about 200 sq mi (population 32,000) appear as eastern outliers of the Puerto Rico mountain mass but were submerged much more recently and hence display marine physical features—notably barrier reefs—superimposed on continental rocks. Anegada, at the extreme northeast of the Greater Antillean chain, is a flat coral reef raised 15 ft above sea level; the other islands have steep slopes and a minimum of flat land, except for St. Croix, south of the main range. Maximum elevation, however, is only 1781 ft (Tortola); rainfall is low and erratic; vegetation, a scanty dry scrub woodland. The Virgin Islands are divided between the United States (St. Croix, St. Thomas, and St. John) and Great Britain (Tortola, Virgin Gorda, Jost Van Dyke, and about 50 other islets, of which six are inhabited).

Lesser Antilles. These numerous islands with a total area of about 2500 sq mi (population about 1,175,000) compose an arc marking the eastern margins of the Caribbean Sea. Built up on a submarine ridge, they stretch 500 mi from the eastern end of the Greater Antilles at Anegada Passage in the north to a 90-mile-wide channel between Grenada and Tobago in the south. Mostly oceanic, the Lesser Antilles are festooned in two, perhaps three, separate chains: an inner ring of volcanic islands; an outer line of predominantly limestone islands to the east, and Barbados, farthest east. The largest, Guadeloupe, is a composite: the western half (Basse Terre) volcanic, the eastern (Grande Terre) limestone. From north to south, the main islands in the two chains are shown in the table.

Island locations of the West Indies. Outstanding inter-island seaways are 1, Yucatan Channel, 2, Windward Passage, 3, Mona Passage; 4, Anegada Passage; 5, Straits of Florida, and 6, Old Bahama Channel (Modified from a map by American Geographical Society)

West Palm Beach, Florida. Most of the islands (Andros Island, 1600 sq mi, is largest) lie on the Great and Little Bahama banks, slightly submerged limestone platforms separated from Florida and Cuba by the shallow Straits of Florida and Old Bahama Channel. Made of calcareous sand of marine origin, and flanked on the east by a coral barrier reef, the Bahamas are flat and low, exceeding 200 ft in height only on Cat Island (400 ft). Precipitation is consequently scanty and unreliable, ranging from 40 in /yr at Nassau, New Providence Island, to 25 in. at Grand Turk. What little soil exists is generally fertile, but the significant produce of land and sea is salt and sponges. All the islands are British; the Turks and Caicos belong to The West Indies federation.

Greater Antilles. These islands comprise the major portion of the West Indies. They represent a partly submerged extension of an ancient Central American mountain chain, which stretches eastward from Hispaniola beneath 75-mile-wide Mona Passage through Puerto Rico and the Virgin Islands to 40-mile-wide Anegada Passage. Westward from Hispaniola the range splits in two. The southern prong runs from the southwestern peninsula of Haiti into the sea, emerges in the Blue Mountains of Jamaica, and reappears to the west in Swan and Roatán islands off northeast Honduras The northern prong runs from Haiti's northwest peninsula across 55-mile-wide Windward Passage, reappearing in the Sierra Maestra of southeastern Cuba and again, further west, in the Cayman Islands (100 sq mi; population 8000; British; in The West Indies federation), low, flat limestone islands with fringing coral reefs, and reaches the shore of Central America north of the Bay of Honduras. Between these two prongs runs Cayman Deep (Bartlett Trough and Oriente Deep), which reaches 22,788 ft below sea level. North of the entire structure, separating Puerto Rico from the Bahama platform, is Puerto Rico Trench, 30,180 ft below sea level in Milwaukee Depth. The extremes of elevation, the intense faulting and folding of most of the strata, gravity anomalies, and earthquakes all attest to the tremendous pressures to which the whole area has been subjected.

Cuba With a length about 760 mi and area of 44,218 sq mi (population 6,410,000) this is the largest of the West Indies. Only its southeastern mountain range, the Sierra Maestra, which reaches 6560 ft at Pico Turquino, links it structurally and physiographically with the other Greater Antilles. The rest of the long, narrow island is essentially a level or gently sloping limestone plateau resembling southern Florida, the Bahamas, and, west of the broad but shallow Yucatan Channel, the peninsula of Yucatan The low-lying plateau is significantly broken only by the Sierra de Trinidad (3700 ft) in the center of the island and by the Sierra de los Órganos at the western end, the latter a broken limestone country of karst topography, with landforms produced by chemical solution rather than by subaerial erosion (*see* KARST TOPOGRAPHY) Lacking a central dividing mountain range, Cuba gets less rainfall (about 50 in annually) than others of the Greater Antilles, but since there is no impediment to the trade winds, the south coast receives as much as the north Only the western end—closest to the usual hurricane track—is much wetter. The typical vegetation is a thorn-

are rare. Subclinical infections are common. *See* ANIMAL VIRUS

The West Nile (WN) virus has properties similar to those of other group B arbor viruses. Only one antigenic type is known. Arbor viruses are a group of arthropod borne animal viruses *See* ANTIGEN

Laboratory diagnosis is made either by the isolation of the virus from the patient's blood by inoculation of the blood into mice, or by serum antibody tests

WN is recurrently epidemic in Israel; in Egypt it is endemic, with immunizing infections, which are mostly silent, occurring in early childhood The infection cycle by which the virus is maintained in nature is thought to be bird → mosquito → bird, with *Culex univittatus* the principal vector. Human infection, as well as that of other vertebrates, occurs seasonally in relation to the presence and density of the infected mosquito population. *See* ARBOR VIRAL ENCEPHALITIDES [J.L.M.]

Bibliography: T. M. Rivers and F. L. Horsfall, Jr., *Viral and Rickettsial Infections of Man*, 3d ed., 1959

Wet cell

A primary cell in which there is a substantial amount of free electrolyte in liquid form. Important examples of wet cells are the Lalande or caustic soda cell the air-depolarized alkaline cell, the air-depolarized sal ammoniac cell, and the Weston standard cell.

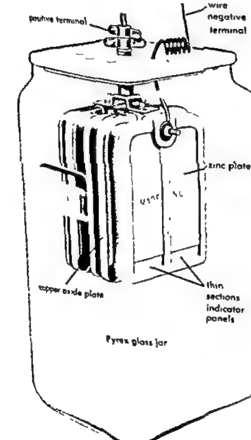


Fig 1 Lalande-Edison copper oxide primary battery

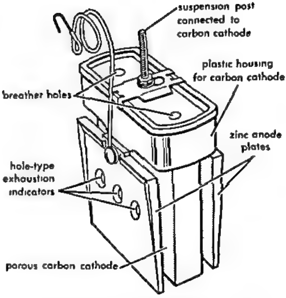


Fig 2 Construction of air-depolarized alkaline cell.

Lalande cell. This cell uses a zinc anode, a cupric oxide cathode and an electrolyte of sodium hydroxide in aqueous solution (caustic soda).

The amalgamated zinc electrodes are cast as flat plates, or as hollow cylinders with thin sections, which corrode through as the copper oxide electrode approaches exhaustion The bright copper color of the cathode can be seen through these openings to warn the user of approaching exhaustion of the cell

The cathode is made by molding cupric oxide into flat plates or hollow cylinders. The oxide is mixed with a binder, pressed, and roasted As it is used, the cupric oxide is partially reduced to metallic copper, which greatly increases the conductivity of the cathode.

The electrolyte is a solution of sodium hydroxide in water of good purity The specific gravity is about 1.21 Normally, the surface of the electrolyte is covered with a layer of oil which retards evaporation of water and absorption of carbon dioxide from the atmosphere.

The anode reaction in the Lalande cell is the oxidation of zinc to form zinc oxide, which dissolves in the electrolyte to form sodium zincate. With sufficient electrolyte, no solid phase forms until the cell is nearly exhausted Then precipitation occurs. Because of this reaction, it is necessary to provide about 8 ml of 1.21 sp gr solution per rated ampere-hour The zinc must be mounted near the top of the electrolyte to prevent

oxide to metallic copper. For this reason, the plate always has the appearance of metallic copper.

The cell potential is 0.95–1.0 volt, but the closed-circuit voltage on normal discharge starts at about 0.65 and decreases slowly to a cutoff voltage of

All British islands belong to The West Indies federation. Martinique and Guadeloupe, with the other French islands, are departments of France. Dutch islands belong to the Netherlands Antilles.

Lesser Antilles

Islands	Maximum elevation, ft	Area, sq mi	Population, thousands
Main volcanic arc			
Saba (Dutch)	2850	5	1
St. Eustatius (Dutch)	1980	8	1
St. Kitts (Br.)	3711	68	36
Nevis (Br.)	3596	50	13
Redonda (Br.)	1000	0.5	0
Montserrat (Br.)	2000	32	14
Guadeloupe, west (Fr.)	4869	264	112
Les Saintes (Fr.)	1036	5	3
Dominica (Br.)	4747	305	66
Martinique (Fr.)	4429	417	260
St. Lucia (Br.)	3145	238	91
St. Vincent (Br.)	4048	133	74
The Grenadines (Br.)	1010	30	15
Grenada (Br.)	2749	120	83
Outliers			
Sombrero (Br.)	40	*	0
Anguilla (Br.)	213	35	7
St. Martin (half Fr., half Dutch)	1350	33	5
St. Barthélemy (Fr.)	990	9	2
Barbuda (Br.)	115	62	1
Antigua (Br.)	1330	108	55
Guadeloupe, east (Fr.)	1300	219	116
La Désirade (Fr.)	930	11	2
Marie Galante (Fr.)	670	58	17
Barbados (Br.)	1104	166	232

* Small fraction of a square mile

Main volcanic arc. In this, islands are spaced at regular intervals, with channels between them nowhere more than 30 mi wide. With hard basaltic cores covered by ash and other igneous material, these islands are the remnants of volcanoes in various stages of decay and subaerial erosion. At the north and south ends of the chain there has been no recent volcanic activity, but from St. Kitts to St. Vincent the cones are still active; in 1902 St. Vincent's Soufrière, Martinique's Mt. Pelée, and Guadeloupe's Soufrière all erupted, causing widespread damage and loss of life. Solfataras, or emissions of hot sulfurous vapors and gases, and fumaroles, or eruptions of steam from subsurface lava evince continued volcanic activity. All extremely mountainous, these islands rise in some cases thousands of feet sheer out of the ocean, the Pitons of St. Lucia being the most spectacular example. There is little level land, and the dense (400/sq mi) populations have occupied all but a small part of even the steepest slopes.

Outliers. By way of contrast, the outlying islands are low-lying and flat or only moderately hilly. They consist of ancient submerged volcanoes capped by sedimentary strata, now slightly tilted. Less fertile and more arid than the inner volcanic arc, they support only a degraded scrub vegetation and are, for the most part, less densely populated

Barbados. Located at the extreme southeast of the chain of outliers, this is a remnant of an ancient continent. Coral limestone escarpments ascend to an elevation of 1104 ft in Mt. Hillaby. Sedimentary strata of continental origin underlie the coral and appear at the surface in the rugged, deeply dissected Scotland District of the northeast. Elsewhere, Barbados' undulating and intensely cultivated surface, is more reminiscent of an English than of a West Indian landscape.

Trinidad and Tobago. These are detached fragments of the Caribbean Coastal Range of the Andes in northern Venezuela, from which Trinidad is separated by the Gulf of Paria. In climate and culture they resemble the Lesser Antilles (they belong to The West Indies federation), but their land forms and vegetation are more continental than insular.

Tobago. A small island (116 sq mi; population about 35,000), Tobago is aligned along a densely wooded ridge which rises to 1800 ft in the northeast and slopes off to the southwest.

Trinidad. This larger island has an area about 1862 sq mi (population about 730,000) and is made up of three mountain ranges with lowlands and swamps between them. The forested northern range, which reaches a height of 3085 ft, is an east-west trending continuation of the Peninsula de Paria of Venezuela to the west, cut off by the Bocas del Dragón and a chain of small islands. The central range bisects the island obliquely from southeast to northwest; the southern range, like the central range only 1000 ft high, forms the southern rim of Trinidad, separated from the Orinoco delta by the Boca de la Sierpe. Between the central and southern ranges are deposits of asphalt, in Pitch Lake, and of oil, which provides four fifths by value of Trinidad's exports.

Other Dutch and Venezuelan islands. Off northern South America several islands appear, like Trinidad and Tobago, as outliers of Andean ranges. Venezuelan Margarita (444 sq mi; population about 70,000), a mountainous island which reaches 4800 ft at its western end, and flat La Tortuga are remnants of the Caribbean Coastal Range. Farther west, the Netherlands Leeward Islands of Aruba, Curaçao, and Bonaire (70, 170 and 108 sq mi, total population 186,000), together with a few small Venezuelan islands, represent an extension of the peninsula of Guajira in Colombia. The Dutch islands are worn-down remnants of crystalline rocks, from which most of an overlying limestone has been eroded. Low-lying and exposed to the sweep of the northeast trades, they are arid (mean annual rainfall at Curaçao, 22 in.) and support only xerophytic, or drought-resistant, vegetation.

[D.L.O.]

West Nile fever

An acute, mosquito-borne virus disease. Mild, atypical, and abortive cases have been reported. The disease occurs in the summer, chiefly in Egypt, Israel, and Africa. Usual signs are fever, rash, and lymphadenopathy; fatality r... ous after-effects

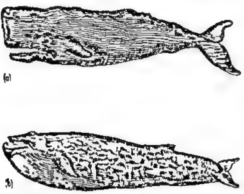
Standard cells are not intended as power sources. They should be used only for comparison of voltages. Ordinary voltmeters put too heavy a drain on the cell for any reliable voltage measurements *See* ELECTROCHEMISTRY; PRIMARY BATTERY. [S EI]

Bibliography: G W. Heise, E. A. Schumacher, and C. R Fisher, The air-depolarized primary cell with caustic alkali electrolyte II, *Trans. Electrochem Soc.*, 92.173, 1947; G. W. Vinal, *Primary Batteries*, 1950

Whale

A marine mammal of the cosmopolitan order Cetacea, highly modified for aquatic life by the development of lateral flukes at the posterior end of the body, loss of the hindlimbs, and modification of [illegible]

which teeth are replaced by baleen (whalebone) plates hanging down from the roof of the mouth through which plankton are strained from the sea.



(a) The sperm whale, *Physeter catodon*, length to 60 ft (b) The blue whale, *Balaenoptera musculus*, length to 109 ft (From P. M Duncan, ed , Cassell's Natural History, Cassell, 1883)

Whales are a valuable source of high-quality oil widely used in soap manufacture and for many other purposes. The commercially important whales are in grave danger of being exterminated, at least as a commercial resource, through virtually unrestricted hunting.

Whales are the largest animals that have ever lived, the biggest being *Balaenoptera* (= *Sibbaldus*) *musculus*, the blue or sulfur-bottomed whale, known to reach a length of 109 ft and a weight of over 150 tons *See* DOLPHIN; PORPOISE [J D B]

Wharf

A structure along a waterfront serving as a berth for ships loading and discharging passengers and cargo A marginal wharf, one parallel to shore, is generally known as a quay or bulkhead. When perpendicular or oblique to shore, a wharf is known as a pier Construction is essentially that required for retaining earth fill or for embankment protection while at the same time providing a landing platform Reinforced-concrete or masonry retaining walls are used as well as the familiar pile-supported platforms with sheet piling bulkheads of steel, timber, and concrete. Movement and settlement of marginal wharves and eventual failure are not uncommon *See* COASTAL ENGINEERING; PIER. [E.J.Q]

Wheat

The most widely grown of all the crop plants in the world is wheat, *Triticum* spp In almost every country, wheat is the main source of flour for bread (Fig. 1). It has been estimated that 2½ bushels of wheat are produced annually for every person in the world. Wheat also is one of the most important of all cereal grains in international trade. The world's largest exporter is Canada Other leading export countries are the United States, Argentina, and Australia Before World War I Russia was the leading exporter of wheat, but exports from the U.S.S.R have declined as domestic needs have increased

The United States leads in wheat production, followed by China and Russia Other leading producers are Canada, India, France, Italy, and Argentina. Production within many smaller countries becomes an important part of their agriculture although it does not contribute greatly to the world's supply of wheat For example, about one-fourth of the 6,000,000 acres of irrigated land in the Nile Valley of Egypt is planted to wheat which is used entirely for domestic consumption

Although wheat has a wide range of adaptation to soils and climate, its greatest concentration of

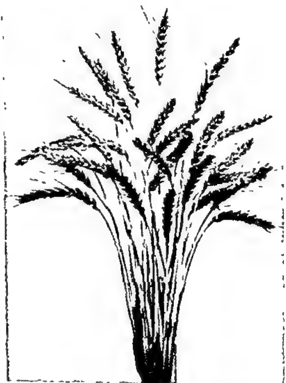


Fig 1 A typical bearded variety of wheat (Iowa

about 0.50. Commercial cells are intended for relatively long continuous service. A 500-amp-hr light-duty cell has an output rating of 1.75 amp at 70°F. A heavy-duty cell of the same ampere-hour rating is capable of continuous output at 6.5–12.0 amp. Unit energy output is about 1.1 kwhr/ft^3 of cell.

At temperatures below 70°F, full output (ampere-hours) can be obtained at a reduced current. Current is reduced by 40% at 40°F, 67% at 20°F, and 83% at 0°F.

Air-depolarized alkaline cell. This cell uses a *zinc anode, a porous carbon cathode exposed to air* on one face, and an alkaline electrolyte. The carbon cathode utilizes atmospheric oxygen. Its zinc anode and alkaline electrolyte are like those of the *Lalande cell, but it has twice the operating voltage* and twice the watt-hour output of an equal-size Lalande cell.

The cathodic reactions have been shown to be:

$$2e + O_2 + H_2O \rightarrow O_2H^- + OH^-$$
$$O_2H^- + OH^- + 2H^+ \rightarrow H_2O_2 + H_2O$$
$$H_2O_2 \rightarrow H_2O + \tfrac{1}{2}O_2$$

The oxygen reacts to form hydrogen peroxide which decomposes readily to water and oxygen. The net amount of oxygen, then, is 0.3 g/amp-hr. At standard temperature and pressure, 210 ml of oxygen are consumed per ampere-hour.

The porous carbon has been reported to have an apparent density of only 0.65 with a porosity of 60%. To function properly, the inner surfaces should be dry. To resist penetration of electrolyte, the pore size must be small and the surfaces partially waterproofed by impregnation with paraffin. The cathode acts as a pump, drawing oxygen from the air. If too great a current is drawn, the pressure in the pore may drop sufficiently to allow electrolyte penetration. This reduces the activity of the carbon and may cause cell failure. Hence it is important that the cell should not be overloaded. For a railway cell, rated at 500 amp-hr, the recommended continuous drain is 2.0 amp at temperatures above 45°F. This is a current density of 3.1 amp/ft^2. Much higher current densities can be obtained with porous carbon electrodes by special design.

The cell open-circuit voltage is 1.46. The railway cell at room temperature (75°F), rated at 500 amp-hrs, will deliver 2 amp at an average of 1.13 volt to a cutoff of 1.05 volt. On 3-amp intermittent signal test, the cell delivers rated capacity of 1.09 volt average at 75°F, 0.98 volt average at 32°F.

Air-depolarized cells are now available in sizes up to 2500 amp-hr. The 2500-amp-hr cell delivers 5.64 kwhr/ft^3 and 0.074 kwhr/lb when discharged at low rates.

Weston standard cell. The Weston cell of 1893 has become the accepted standard of electromotive force. The *Weston normal or saturated cadmium* cell has an electromotive force (emf) of 1.01864 absolute volts at 20°C. When made of purified materials, cells having the same emf to within a few

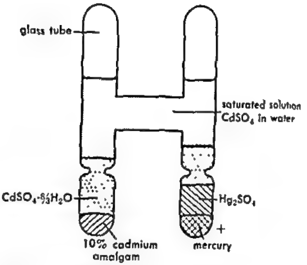


Fig. 3. Schematic of Weston saturated standard cell

emf very well. Reference standards in daily use for many years are remarkably constant. These standard cells are made with materials of spectroscopic purity and are maintained under diffuse light in a thermostatically controlled oil bath in a room maintained at 25°C and 60% relative humidity. *See* VOLTAGE MEASUREMENT.

The cell uses a 2-phase amalgam of cadmium as the anode. For a 10% amalgam, 1 part of cadmium by weight and 9 parts of mercury are required. These materials may be combined either by heating them together or by electrolytic deposition of cadmium into mercury. At ordinary temperatures a liquid phase is in equilibrium with a solid phase. This gives a very stable potential which depends only on the temperature.

The cathode is mercurous sulfate, Hg_2SO_4, in contact with mercury.

The electrolyte is a saturated solution of cadmium sulfate in equilibrium with the solid phase $CdSO_4 \cdot \frac{8}{3}H_2O$. In some cells, sulfuric acid is added to prevent hydrolysis of mercurous sulfate.

The cell is usually made of glass in the form of an H, as in Fig. 3. *Platinum-wire leads are sealed* in the base of each arm. Mercury, carefully purified, is placed at the bottom of one arm and a 10% cadmium amalgam is placed, while warm and in a single phase, in the other arm. When the amalgam has cooled and separated into two phases, crystals of mercurous sulfate are placed above the mercury and crystals of cadmium sulfate are placed above the amalgam. A saturated solution of cadmium sulfate is then added to about 2–3 mm above the crossbar, and the cell is hermetically sealed.

The best-saturated cells of this type may be *measured to the ten-millionth part of a volt* at specified temperatures which must be known to within 0.01°C.

The saturated Weston cell has a relatively large *temperature coefficient of emf. For portable use*, it is general practice to use a cell with an unsaturated electrolyte. This has a temperature coefficient which is only one fourth as great as that of the satu-

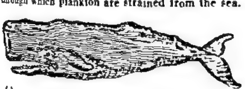
Standard cells are not intended as power sources. They should be used only for comparison of voltages. Ordinary voltmeters put too heavy a drain on the cell for any reliable voltage measurements. *See* ELECTROCHEMISTRY; PRIMARY BATTERY. [S.F.L.]

Bibliography: G. W. Heise, E. A. Schumacher, and C. R. Fisher, The air-depolarized primary cell with caustic alkali electrolyte II, *Trans. Electrochem. Soc.*, 92:173, 1947; G. W. Vinal, *Primary Batteries*, 1950

Whale

A marine mammal of the cosmopolitan order Cetacea, highly modified for aquatic life by the development of lateral flukes at the posterior end of the body, loss of the hindlimbs, and modification of

which teeth are replaced by baleen (whalebone) plates hanging down from the roof of the mouth through which plankton are strained from the sea.



(a)

(b)

(a) The sperm whale, *Physeter catodon*, length to 60 ft (b) The blue whale, *Balaenoptera musculus*, length to 109 ft (From P. M. Duncan, ed., *Cassell's Natural History*, Cassell, 1883)

Whales are a valuable source of high-quality oil widely used in soap manufacture and for many other purposes. The commercially important whales are in grave danger of being exterminated, at least as a commercial resource, through virtually unrestricted hunting.

Whales are the largest animals that have ever lived, the biggest being *Balaenoptera* (= *Sibbaldus*) *musculus*, the blue or sulfur-bottomed whale, known to reach a length of 109 ft and a weight of over 150 tons. *See* DOLPHIN; PORPOISE. [J.D.B.]

Wharf

A structure along a waterfront serving as a berth for ships loading and discharging passengers and cargo. A marginal wharf, one parallel to shore, is generally known as a quay or bulkhead. When perpendicular or oblique to shore, a wharf is known as a pier. Construction is essentially that required for retaining earth fill or for embankment protection while at the same time providing a landing

Reinforced concrete or masonry retain-

re used as well as

ported platforms with sheet piling bulkheads of steel, timber, and concrete. Movement and settlement of marginal wharves and eventual failure are not uncommon. *See* COASTAL ENGINEERING; PIER. [F.J.Q.]

Wheat

The most widely grown of all the crop plants in the world is wheat, *Triticum* spp. In almost every country, wheat is the main source of flour for bread (Fig. 1). It has been estimated that 2½ bushels of wheat are produced annually for every person in the world. Wheat also is one of the most important of all cereal grains in international trade. The world's largest exporter is Canada. Other leading export countries are the United States, Argentina, and Australia. Before World War I Russia was the leading exporter of wheat, but exports from the U.S.S.R. have declined as domestic needs have increased.

The United States leads in wheat production, followed by China and Russia. Other leading producers are Canada, India, France, Italy, and Argentina. Production within many smaller countries becomes an important part of their agriculture although it does not contribute greatly to the world's supply of wheat. For example, about one-fourth of the 6,000,000 acres of irrigated land in the Nile Valley of Egypt is planted to wheat which is used entirely for domestic consumption.

Although wheat has a wide range of adaptation to soils and climate, its greatest concentration of

Fig. 1 A typical bearded variety of wheat. (Iowa

production is in those areas where the corn crop cannot be successfully grown because of either limited rainfall or a short frost-free growing season. It is because of a too short growing season for corn that the prairie provinces of central Canada and the states of North Dakota, South Dakota, and Montana rank high in wheat acreage. In these areas, wheat is planted in the spring. The central plains states, Kansas, Oklahoma, Nebraska, and Texas, too dry for corn, grow large acreages of wheat planted in the fall to utilize best the more favorable soil moisture conditions in the cooler seasons of the year. The Corn Belt states of Illinois, Indiana, Ohio, and southern Michigan also are important in wheat production, largely growing fall-sown types In the far west, Washington, Oregon, and Idaho grow wheat under low rainfall The Palouse valley in Washington is famous for its intensity of production. The 10-year average farm value of wheat in the United States for the period 1945–1954 was over $2,500,000,000

Market classes. Wheat used for milling purposes is grouped into seven commercial classes. These include (1) hard red spring, used for bread flour, (2) durum and (3) red durum, both used for macaroni and similar products, (4) hard red winter, used for bread flour, (5) soft red winter, used largely for cake and pastry flour, (6) white wheat, which may be either hard or soft, used for bread or pastry flour, and (7) mixed wheat. Among the market classes grown in the United States, the hard red winter type comprises over 50%, hard red spring about 20%, soft red winter about 10%, white wheat less than 10%, and durums about 2% of the total production. The distribution of the four major market classes is shown in Fig 2.

Origin and description. *The culture of wheat* as a bread cereal reaches back to prehistoric times Carbonized grains of wheat have been found in Iraq, tracing to an origin nearly 7000 years ago It is quite possible that wheat was being grown even before this date. Several world centers have been presumed to be the evolutionary centers of origin for wheats, including central Asia, the Near East, the Mediterranean area, and Abyssinia

The genus *Triticum* is composed of many species that can be grouped according to their chromosome numbers of 7, 14, and 21 pairs (*see* GENETICS) Species with 7 pairs of chromosomes include *T. aegilopoides*, a wild hulled wheat, and *T. monococcum*, a cultivated hulled wheat grown only to a limited extent. The species with 14 pairs of chromosomes include *T. dicoccoides*, a wild hulled emmer, *T. dicoccum*, the cultivated emmer, *T. durum*, the cultivated durums, *T. persicum*, Polish wheat, *T. turgidum*, Rivet wheat, *T. polonicum*, also called Polish wheat, and *T. timopheevi*. The last four species in this group are not extensively grown. The species in the 21-chromosome group include *T. vulgare* the most widely grown of all wheats, *T. compactum*, the club wheats grown largely in the far west, and *T. spelta*, the hulled spelt wheat grown to a limited extent for livestock feed. Other species in this group, including *T. sphaerococcum* and *T. macha*, are not cultivated.

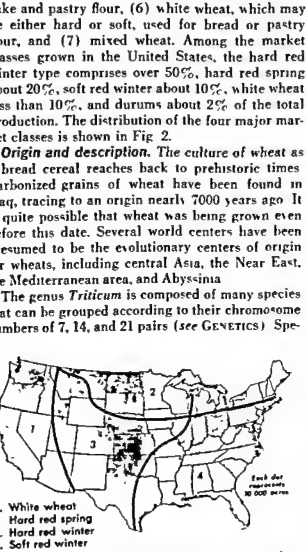


Fig. 2. Areas of production of the four major market classes of wheat in the United States

Because of the great importance of wheat as a cultivated crop, many studies have been made in an attempt to learn the genetic relationships that exist among species of the same chromosome groups and among species that differ in chromosome number. More recently studies have been made to determine the origin of cultivated wheats as related to other genera. Briefly, these studies show that species with the same chromosome number generally can be freely intercrossed, and the hybrids are fertile. Crosses also can be made between the 7- and 14-chromosome species and between the 14- and 21-chromosome species. Such species crosses exhibit considerable sterility due to lack of complete chromosome pairing. For example, in crosses between *T. durum* (14 pairs) and *T. vulgare* (21 pairs), 14 of the 21 *T. vulgare* chromosomes pair nearly normally with the 14 chromosomes from *T. durum* The remaining 7 chromosomes of *T. vulgare* remain unpaired, thus causing sterility. A few progenies in later generations from interspecific hybrids contain 21 pairs of chromosomes and are fully fertile Hybrids between 14- and 21-chromosome species have been successfully used to transfer desired characters for disease resistance from durum and emmer to common wheats. *See* BREEDING (PLANT).

Certain studies to determine the origin of the 21 chromosomes in *T. vulgare* suggest that 7 of them originated from the forage grass, *Agropyron*, 7 from the genus *Aegilops*, and 7 from the genus *Triticum* In this proposed origin, hybrids between *Agropyron* and a 7-chromosome *Triticum* produced the 14-chromosome species of *Triticum* and this hybrid, in turn, when crossed to *Aegilops* produced common wheat Thus, the genetic origin of wheat is no doubt a complex one This evolutionary process may have occurred over a long period of time before primitive man began to use this grain as a human food

All cultivated species of wheat are self pollinated, with less than 1% of natural crossing (*see* REPRODUCTION, PLANT) Plants vary in height from 2 to 5 ft, depending upon variety, soil fertility, and moisture supply Most varieties are hollow stemmed and may produce several tillers (shoots) from the lower nodes or crowns *See* STEM (BOTANY). The inflorescence (flower ' ister) is a spike.

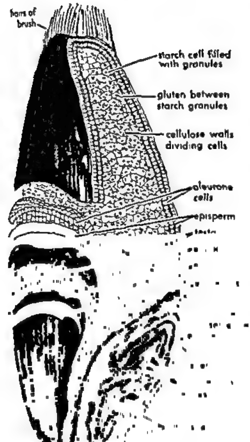


Fig. 3 Longitudinal section of a grain of wheat. (From L. E. Booher and I. T. Behan, Nutrient analyses of U.S. food supplies, J. Nutrition, 39(4) 495–515, 1949)

with spikelets alternately arranged sessile to a zigzag rachis. *See* FLOWER (BOTANY), GRASS CROPS. Spikelets (except einkorn and emmer) contain 2–3 fertile florets, each enclosed in a lemma and palea. The caryopsis (grain) is easily freed from the glumes in threshing (Fig. 3). The lemma in some varieties is awnless, in others short awned or fully awned. The grain may be red or white, and it may be hard, as in hard red spring and hard red winter market classes, or soft, as in soft red winter and certain white-wheat market classes. Durum wheats usually have larger grains than bread wheats. *See* FRUIT (BOTANY).

Varieties. Because of the importance of wheat, plant breeders in many countries have bred new varieties either by pure-line selection (in the pre-Mendelian era) from existing mixed types, or by modern methods of hybridization between parental varieties carefully chosen to give desired combinations of characters to their progenies. Many thousand varieties are available in the world and nearly 200 recognized varieties are grown to greater or less extent in the United States. Wheat varieties must yield flour possessing exacting milling and baking qualities to meet the needs for commercial purposes or in the home. Often wheats of different varieties and market classes must be blended to produce a flour of desired gluten quality for varied uses. Wheat varieties also must have desired agronomic characters such as high yield, winter hardiness in the winter-wheat areas, earliness in the northern states, resistance to lodging, and resistance to destructive diseases and insect pests. Development of wheat varieties that meet all these requirements is difficult. Successful attainment of these objectives has been possible by teamwork among breeders, pathologists, entomologists, and cereal chemists working at many research centers. The pedigree of five spring wheat varieties illustrates the complex parentage in the origin of these new wheats (Fig. 4).

Cultural practices. In the spring-wheat and the Great Plains winter-wheat areas of the United States, wheat generally is planted after another small grain crop in the rotation. In the Corn Belt states, winter wheat often is planted after soybeans, and in the arid regions of the far west and in the western Great Plains under dry land farming methods, wheat generally is planted after the land has been fallowed to build up moisture reserves. Spring wheats are planted as early as a good seedbed can be prepared, and winter wheats from late October in Texas to mid-September in the northern limits of its production. Often the Hessian fly-free date is the determining factor in planting winter wheat (*see* DIPTERA). Wheat is commonly planted at the rate of 1¼–1½ bushels per acre at a depth of 1–2 in., depending on soil moisture.

Wheat is generally harvested with a combine direct from the standing crop when moisture content is 13% or less, but it may first be windrowed and then combined when the grain is sufficiently dry to store safely. *See* AGRICULTURAL MACHINERY.

[I.J.J.]

WHEAT DISEASES

It is estimated that in the United States more than 76,000,000 bushels of wheat are destroyed annually by diseases. Some diseases, such as stem rust and scab, become epidemic only in certain years; others, such as root rots, are prevalent every year. When diseases like leaf and stem rusts, smuts, and scab become epidemic, they may destroy 25 to 50% of the potential crop, both locally and regionally. Thus in 1935 and 1937 [illegible] wheat cr[illegible] the prair[illegible] durum w[illegible] of wheat stem rust in Minnesota and the Dakotas.

Diseases of wheat are caused chiefly by fungi, bacteria, nematodes and viruses (see BACTERIA; FUNGI, NEMATODA, PLANT VIRUS). The more important ones are stem rust (*Puccinia graminis tritici*), leaf rust (*P. condita*), bunt (*Tilletia* spp.), scab or fusarial head blight, caused by many species of *Fusarium*, particularly *F. roseum* (Fig. 5), and seedling blight and root rot, caused by many species of fungi, especially *Fusarium* [illegible]

production is in those areas where the corn crop cannot be successfully grown because of either limited rainfall or a short frost-free growing season. It is because of a too short growing season for corn that the prairie provinces of central Canada and the states of North Dakota, South Dakota, and Montana rank high in wheat acreage. In these areas, wheat is planted in the spring. The central plains states, Kansas, Oklahoma, Nebraska, and Texas, too dry for corn, grow large acreages of wheat planted in the fall to utilize best the more favorable soil moisture conditions in the cooler seasons of the year. The Corn Belt states of Illinois, Indiana, Ohio, and southern Michigan also are important in wheat production, largely growing fall-sown types. In the far west, Washington, Oregon, and Idaho grow wheat under low rainfall. The Palouse valley in Washington is famous for its intensity of production. The 10-year average farm value of wheat in the United States for the period 1945–1954 was over $2,500,000,000.

Market classes. Wheat used for milling purposes is grouped into seven commercial classes. These include (1) hard red spring, used for bread flour, (2) durum and (3) red durum, both used for macaroni and similar products, (4) hard red winter, used for bread flour, (5) soft red winter, used largely for cake and pastry flour, (6) white wheat, which may be either hard or soft, used for bread or pastry flour, and (7) mixed wheat. Among the market classes grown in the United States, the hard red winter type comprises over 50%, hard red spring about 20%, soft red winter about 10%, white wheat less than 10%, and durums about 2% of the total production. The distribution of the four major market classes is shown in Fig 2

Origin and description. The culture of wheat as a bread cereal reaches back to prehistoric times Carbonized grains of wheat have been found in Iraq, tracing to an origin nearly 7000 years ago. It is quite possible that wheat was being grown even before this date. Several world centers have been presumed to be the evolutionary centers of origin for wheats, including central Asia, the Near East, the Mediterranean area, and Abyssinia.

The genus *Triticum* is composed of many species

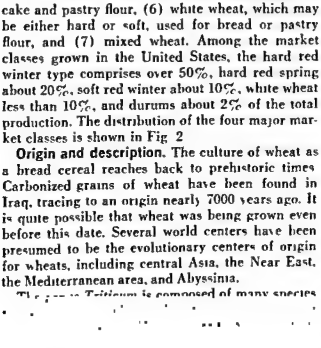


Fig. 2. Areas of production of the four major market classes of wheat in the United States.

cies with 7 pairs of chromosomes include *T. aegilopoides*, a wild hulled wheat, and *T. monococcum*, a cultivated hulled wheat grown only to a limited extent. The species with 14 pairs of chromosomes include *T. dicoccoides*, a wild hulled emmer, *T. dicoccum*, the cultivated emmer, *T. durum*, the cultivated durums, *T. persicum*, Polish wheat, *T. turgidum*, Rivet wheat, *T. polonicum*, also called Polish wheat, and *T. timopheevi*. The last four species in this group are not extensively grown. The species in the 21-chromosome group include *T. vulgare* the most widely grown of all wheats, *T. compactum*, the club wheats grown largely in the far west, and *T. spelta*, the hulled spelt wheat grown to a limited extent for livestock feed. Other species in this group, including *T. sphaerococcum* and *T. macha*, are not cultivated.

Because of the great importance of wheat as a cultivated crop, many studies have been made in an attempt to learn the genetic relationships that exist among species of the same chromosome groups and among species that differ in chromosome number. More recently studies have been made to determine the origin of cultivated wheats as related to other genera. Briefly, these studies show that species with the same chromosome number generally can be freely intercrossed, and the hybrids are fertile. Crosses also can be made between the 7- and 14-chromosome species and between the 14- and 21-chromosome species. Such species crosses exhibit considerable sterility due to lack of complete chromosome pairing For example, in crosses between *T durum* (14 pairs) and *T. vulgare* (21 pairs), 14 of the 21 *T. vulgare* chromosomes pair nearly normally with the 14 chromosomes from *T. durum*. The remaining 7 chromosomes of *T. vulgare* remain unpaired, thus causing sterility. A few progenies in later generations from interspecific hybrids contain 21 pairs of chromosomes and are fully fertile. Hybrids between 14- and 21-chromosome species have been successfully used to transfer desired characters for disease resistance from durum and emmer to common wheats *See* BREEDING (PLANT).

Certain studies to determine the origin of the 21 chromosomes in *T. vulgare* suggest that 7 of them originated from the forage grass, *Agropyron*, 7 from the genus *Aegilops*, and 7 from the genus *Triticum*. In this proposed origin, hybrids between *Agropyron* and a 7-chromosome *Triticum* produced the 14-chromosome species of *Triticum* and this hybrid, in turn, when crossed to *Aegilops* produced common wheat Thus, the genetic origin of wheat is no doubt a complex one This evolutionary process may have occurred over a long period of time before primitive man began to use this grain as a human food

All cultivated species of wheat are self pollinated, with less than 1% of natural crossing (*see* REPRODUCTION, PLANT) Plants vary in height from 2 to 5 ft, depending upon variety, soil fertility, and moisture supply Most varieties are hollow stemmed and may produce several tillers (shoots) from the lower nodes or crowns. *See* STEM (BOTANY). The inflorescence (flower cluster) is a spike.

thosporium sativum, Ophiobolus graminis, Lepto- [illegible]

[illegible] consists of more than 250 [illegible]

processing operations Bunt and smut spores coat surfaces of threshed grain and this necessitates washing before milling; the cost of this extra operation is reflected in lower prices offered the farmer for smutted grain. Fusarial head blight is most prevalent in the Corn Belt where the temperature and relative humidity are high during blossoming period Infected kernels contain substances toxic to man and certain animals.

Root rots of wheat are ubiquitous diseases They are debilitating and insidious, but usually inconspicuous. Most of the root-rotting pathogens live in or on the seed, in soil, and on dead plants. They attack all underground parts of plants, blight the seedlings, rot the roots, and kill adult plants prematurely

Wheat kernels are frequently discolored by fungi and bacteria, which often serve as the forerunners of infection for seedling blights and root rots, and [illegible]

[illegible] nels after harvest and make the grains unsuitable for milling and baking

Control measures. Although diseases of wheat cannot be prevented completely, they can be reduced substantially by good agricultural practices. The use of sound seed, treated with fungicides, eliminates the pathogens from the seed and protects them from soil borne organisms (*see* FUNGISTAT AND FUNGICIDE). A good cropping sequence, coupled with the use of disease-resistant varieties, is highly recommended. Although most varieties may be only temporarily resistant, this greatly reduces the danger of destructive epidemics, particularly of stem rust and bunt.

[J J C.]

Bibliography: See AGRICULTURAL SCIENCE (PLANT), PLANT DISEASE.

WHEAT PROCESSING

The milling process breaks open the wheat kernel and reduces the particles formed so as to separate the outer and inner portions of the kernel Bran and germ are almost completely separated from the white interior portions of the kernel in the milling of refined flour Milling of wheat was brought about through efforts to remove impurities and foreign material and after the discovery that the inside of the wheat kernel tastes better than the outside Grinding wheat also aids mastication, and milling is necessary to produce flour for use in fabricated foods, such as bread

Sifted meal keeps better than unsifted and produces a whiter product Traditionally, white is associated with purity

When the entire kernel is ground and no separations are made, the product is whole-wheat flour.

Flour milling has become one of the principal industries in terms of value of product processed and of value added to the product during processing. Milling is a mechanical process, therefore, the number of persons employed per unit investment is less than for most industries

The following steps are involved in milling wheat: wheat selection, blending, cleaning, conditioning or tempering, breaking, bolting, sieving, purifying, reducing, classifying, and some combination of maturing, bleaching, and diastatic or enriching treatment of the flour.

Cleaning is done with machines that use sieves and air currents to separate dirt, chaff, and foreign seeds from wheat Other devices are available to remove metal or stone contamination and to wash wheat before it is milled Water and often heat are added in the process called tempering or conditioning before the actual milling operation begins, to prevent the bran from fragmenting unduly during milling and to maintain proper milling conditions.

The initial step in actual milling, after the cleaning and tempering, is cracking wheat to small pieces between steel rolls, which are corrugated with small grooves This is the breaking system The pieces of broken kernel of various sizes and dimensions are separated by rotating sieves and by purifiers which combine sieves and air current to separate particles on the basis of size and density. The stocks thus produced are directed to either additional break rolls or smooth steel reduction rolls The flow of material in a typical flour mill is as illustrated

Note that the flour produced comes from many different machines in the milling process These

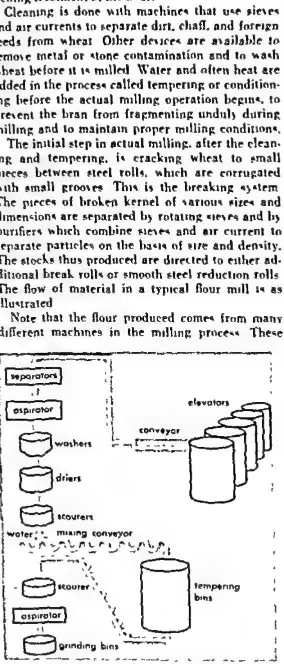


Fig 6 Flow sheet of wheat cleaning (From M E Parker, E H Harvey, and E S Stateler, *Elements of Food Engineering*, vol 1, Reinhold, 1952)

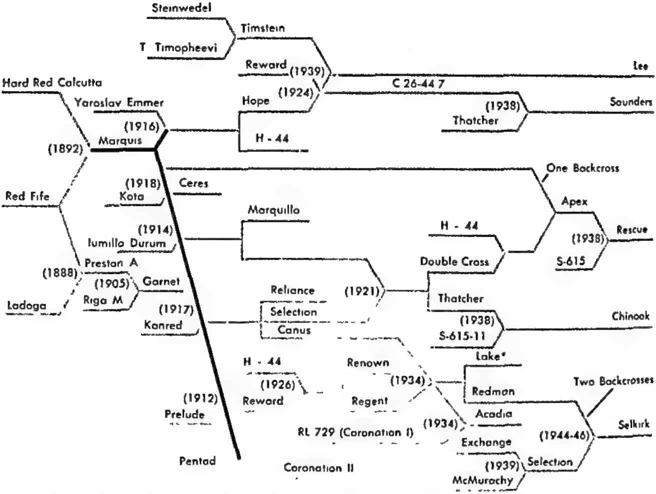


Fig. 4 Modern wheat varieties may have a complex parentage as shown for five spring wheats. Note that 14-chromosome emmer and durum wheat have contributed to these pedigrees

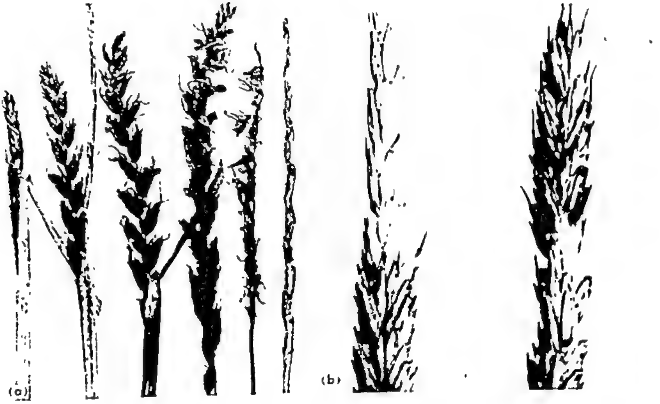


Fig 5 (a) Loose smut, *Ustilago nuda*, of wheat (b) Fusarial head blight (*Minn Agr. Exp. Sta*)

blades rotating in opposite directions. The flour particles are projected at high velocities against hard blades or pins or the surface of the machine where the force of impact causes the particles to break into smaller pieces; correct use of air streams can help increase particle impact force.

Wheat flour contains a low-protein, high-starch fraction and a high protein, low-starch fraction; thus a protein shift can be accomplished by the separation of these two fractions. Particle size separations at cut points of 15–40 μ will separate fractions consisting principally of starch granules. Below 15 μ are principally broken starch granules and much of the protein matrix. Fine grinding makes the "shift" more complete because free starch granules have a particle size range from about 1 to 50 μ, while protein particles range from about 1 to 5 μ. Particles in these size ranges can be separated by air classification where they are actuated by a strong centrifugal force, produced by a rapidly rotating disk. As the particles are thrown toward the perimeter, they are met by strong air currents moving in the opposite direction. The air stream composes a drag on the moving particles, first slowing the movement of the smaller, lighter material, eventually bringing them to a stop, and then reversing the direction of movement. Adjustment of either the rotating force or the velocity of the air stream makes possible a selection in the size of the particles separated.

The utility of this process is that from the grinding of wheat flour, fractions can be obtained which are high in protein or high in starch. High-protein fractions are utilized in supplementing bakers' flour, while low protein flour is used for pastry purposes, and both the protein and starch fractions have industrial uses. [J.A SH.]

Bibliography: E S Miller, *Studies in Practical Milling*, 1941.

Wheatstone bridge

A device used to measure the electrical resistance of an unknown resistor by comparing it with a known standard resistance. This method was first described by S H Christie in 1833, only 7 years after Georg S Ohm discovered the relationship between voltage and current. Since 1843 when Sir Charles Wheatstone called attention to Christie's work, Wheatstone's name has been associated with this network. *See* RESISTANCE MEASUREMENT.

The Wheatstone bridge network consists of four resistors R_{AB}, R_{BC}, R_{CD}, and R_{AD} interconnected as shown in Fig 1 to form the bridge. A detector G, having an internal resistance R_G, is connected between the B and D bridge points; and a power supply, having an open circuit voltage E and internal resistance R_B, is connected between the A and C bridge points. *See* BRIDGE CIRCUIT.

Application of Ohm's and Kirchhoff's laws to this network results in the following expression for the detector current

$$I_G = \frac{I_B(R_{BC}R_{AD} - R_{AB}R_{CD})}{R_G(R_{AB} + R_{BC} + R_{CD} + R_{AD}) + (R_{BC} + R_{CD})(R_{AB} + R_{AD})} \quad (1)$$

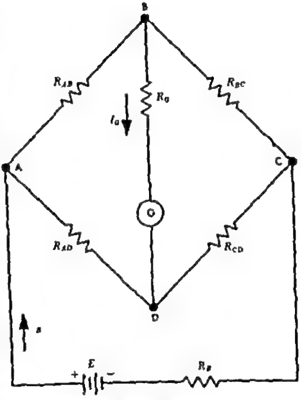


Fig 1 Wheatstone bridge circuit

in which

$$I_B = \frac{E}{R_B + \dfrac{(R_{AB} + R_{BC})(R_{AD} + R_{CD})}{R_{AB} + R_{BC} + R_{CD} + R_{AD}}} \quad (2)$$

It is apparent, from Eq (1), that if the network is adjusted so that

$$R_{BC}R_{AD} - R_{AB}R_{CD} = 0 \quad (3)$$

the detector current will be zero and this adjustment will be independent of the supply voltage, the supply resistance, and the detector resistance. Thus, when the bridge is balanced

$$R_{BC}R_{AD} = R_{AB}R_{CD} \quad (4)$$

and, if it is assumed that the unknown resistance is the one in the CD arm of the bridge, then

$$R_{CD} = \left(\frac{R_{BC}}{R_{AB}}\right) \times R_{AD} \quad (5)$$

Three methods of adjustment to achieve this condition are possible when the circuit is used as a ratio arm bridge: (1) use of a fixed ratio R_{BC}/R_{AB} and a continuously adjustable standard R_{AD}, (2) use of a continuously adjustable ratio and a fixed standard, and (3) a combination of the foregoing with the ratio usually adjustable in discrete steps of decade values. The first method provides a linear calibration of unknown versus standard resistance but is limited in resistance range to the adjustable range of the standard. The second method provides

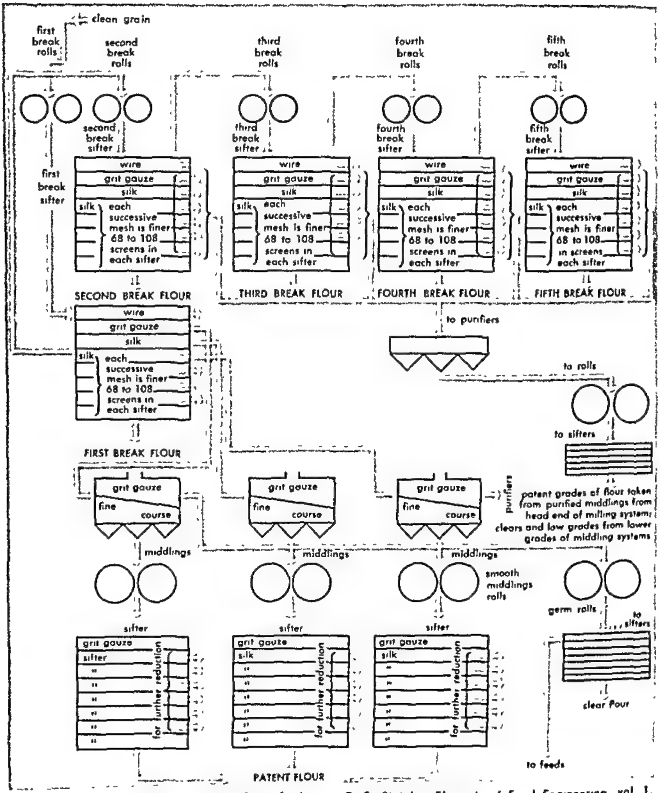


Fig 7. Diagrammatic sketch showing flow of wheat milled into flour. (From M. E. Parker, E H. Harvey, and E S Stateler, *Elements of Food Engineering*, vol 1, Reinhold, 1952)

flour streams can differ in quality. It is by combining flour streams that flour of different grades is produced. "Patent" grade is that flour portion freest from bran and germ contamination The portions of the total flour not included in the patent grade are called "clears." A "straight" grade represents the total flour yield. The feed portion is about equally divided between bran and shorts.

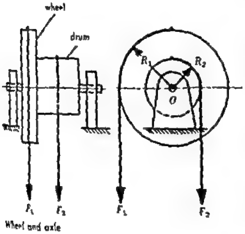
Flours are milled for many purposes such as bread, pastries, cakes, cookies, or macaroni. This necessitates wheat selection and different processing, to some extent, although the basic principles are not changed Flours are manufactured from hard, soft, or durum wheat or combinations of these The yield of 100 lb of wheat is usually 72 lb of flour and 28 lb of feed

One of the newer developments in wheat processing is fine grinding and air classification of flour particles. Through this process, the flour produced by conventional milling systems can be more finely ground and fractionated into products having different compositions and properties. Grinding is accomplished in high speed centrifugal machines sometimes with centrifugal rotors having pins or

For this reason, bridge ratio arms are often adjusted for a specified error in ratio, maintaining only a nominal resistance value. [C.E.A.]

Wheel and axle

A wheel and its axle or, more generally, two wheels with different diameters or a wheel and drum, as in a windlass, rigidly connected together so that they



Wheel and axle

about the axis of rotation equals zero. Where flexible members, such as ropes, have been firmly attached to a wheel and drum (as illustrated) and the machine is mounted on frictionless bearings

$$F_1R_1 - F_2R_2 = 0$$

If F_1 is the input and F_2 the output, the mechanical advantage is

$$\text{MA} = \frac{F_2}{F_1} = \frac{R_1}{R_2}$$

If the bearings are not frictionless, a friction torque T_f will oppose the rotation. For the case with F_1 the input, the friction torque will act in the same direction as the load or output torque F_2R_2, and the equilibrium equation becomes

$$F_1R_1 - F_2R_2 - T_f = 0$$

and

$$F_2 = \frac{F_1R_1 - T_f}{R_2}$$

Because all members are rigidly connected, all torques act through the same angular displacement, the efficiency is then

$$\eta = \frac{F_2R_2}{F_1R_1} \times 100 = \frac{F_1R_1 - T_f}{F_1R_1} \times 100$$
$$= 1 - \frac{T_f}{F_1R_1} \times 100$$

The main difference between the lever and the wheel and axle is that the wheel and axle permits the forces to operate through a much greater distance. For example, in the illustration the wheel and drum could be allowed to rotate any number of revolutions if the ropes were wrapped the required number of times around each before they were attached.

The wheel and axle is seldom encountered in the exact form shown in the illustration, but the principle of the summation of torques about the axis of rotation equalling zero finds widespread applications in machines such as the belt drive, chain drive, friction drive, and gear drive. See SIMPLE MACHINE. [R.M.PH.]

Wheel base

The distance in the direction of travel from front to rear wheels of a vehicle, measured between centers of ground contact under each wheel. For a vehicle with two rear axles, the rear measuring point is on the ground midway between rear axles. Tread of a vehicle is the distance perpendicular to the direction of travel between front wheels, or between rear wheels, measured from centers of ground contact.

Wheel bug

A large, gray member, *Arilus cristatus*, of the family Reduviidae, order Hemiptera, characterized by a semicircular crest on the pronotum, which is notched to resemble a cogwheel. These bugs are capable of inflicting a very painful bite. Occasionally the victim of such a bite may show serious allergic reactions. The wheel bug is relatively common and is predaceous upon other insects. It is related to the kissing bugs. See HEMIPTERA. [J.D.H.]

Whelk

A common name used widely, and apparently indiscriminately in the United States to identify any of a variety of marine gastropods having whorled shells. Originally the name appears to have been applied to certain gastropods of Europe and later to American snails that are sometimes rather distantly related. All whelks have whorled shells and are carnivorous. The name whelk is limited to two families, the Buccinidae and Neptuneidae.

Busycon carica, known also as *Fulgur carica*, and with a variety of common names including knobbed whelk and knobbed pear conch, is one of the largest and best known. This is the largest gastropod found north of Cape Hatteras, growing in favorable waters to a maximum length of 9 in. The pear-shaped shell is large and thick, with six whorls. Its color is yellowish gray, the interior orange red. Those from sheltered waters are not quite as large as specimens from the open beach. It ranges from Cape Cod to Texas. This powerful carnivore bores holes in the largest of oysters, or other bivalves, and sucks their soft parts through the opening, in the manner of the small oyster drills.

The eggs of this whelk are laid in a long parchmentlike ribbon containing as many as 100 egg cases. Such an egg ribbon may be a yard long, and may be laid at any time during the warm months in the Atlantic

a wide range, since the ratio is easily adjustable from zero to infinity, but results in a nonlinear scale, highly expanded for low resistances and greatly compressed for high resistances. The third method, using a ratio adjustable in several decade steps and a resistance standard of three to five decades, provides a wide range and linear calibration and is the most practical of the combinations for general use with reasonable accuracy.

If the circuit is considered as a product arm bridge,

$$R_{CD} = \frac{R_{BC}R_{AD}}{R_{AB}} \tag{6}$$

it is seen that the conductance G_{CD} of the unknown resistance R_{CD} is measured directly in terms of the adjustable standard R_{AB} since

$$G_{CD} = \frac{1}{R_{CD}} = \frac{R_{AB}}{R_{BC}R_{AD}} \tag{7}$$

Sensitivity. The sensitivity of the bridge assembly (battery, bridge, and detector) is of interest for two purposes (1) to determine the required detector sensitivity for a given deviation in the unknown resistance, or (2) to determine the change in resistance which can be measured using a detector of a stated sensitivity. The precision of balance is affected by the detector sensitivity, the detector resistance, the ohmic value of the bridge resistors, the bridge supply voltage, and the bridge supply resistance, which except for special cases can be neglected.

If the detector circuit BD (Fig. 1) is opened, the open-circuit voltage existing between points B and D is

$$e = E_{BC} - E_{CD}$$

or

$$e = E\left(\frac{R_{BC}}{R_{AB} + R_{BC}} - \frac{R_{CD} - \Delta R_{CD}}{R_{CD} - \Delta R_{CD} + R_{AD}}\right) \tag{8}$$

where ΔR_{CD} is a small, incremental change in the unknown resistance R_{CD}. To a close approximation, this open-circuit voltage due to a small bridge unbalance can be expressed in terms of the bridge ratio, the fractional change in the unknown resistance, and the applied voltage as

$$e = E\frac{r}{(r+1)^2}\left(\frac{\Delta R_{CD}}{R_{CD}}\right) \tag{9}$$

where $r = R_{BC}/R_{AB}$.

Sensitivity can also be expressed in terms of the unbalance voltage per volt applied to the bridge

$$e' = \frac{e}{E} = \frac{r}{(r+1)^2}\left(\frac{\Delta R_{CD}}{R_{CD}}\right) \tag{10}$$

If the detector circuit is closed, a current flows through the detector. Neglecting the battery resistance, this current can be calculated from Thévenin's theorem and Fig. 2 as

$$I_G = \frac{e}{R_G + \dfrac{R_{AB}R_{BC}}{R_{AB} + R_{BC}} + \dfrac{R_{AD}R_{CD}}{R_{AD} + R_{CD}}} \tag{11}$$

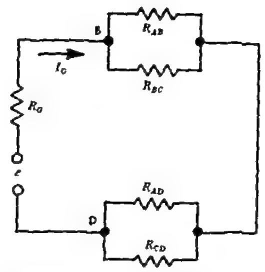


Fig 2 Equivalent circuit of unbalanced bridge

$$\text{or} \quad I_G = \frac{E\left(\dfrac{\Delta R_{CD}}{R_{CD}}\right)}{\dfrac{R_G}{\frac{r}{(r+1)^2}} + R_{AB} + R_{BC} + R_{CD} + R_{AD}} \tag{12}$$

and in terms of unbalance current per volt applied to the bridge

$$I_G' = \frac{I_G}{E} = \frac{\left(\dfrac{\Delta R_{CD}}{R_{CD}}\right)}{\dfrac{R_G}{\frac{r}{(r+1)^2}} + R_{AB} + R_{BC} + R_{CD} + R_{AD}} \tag{13}$$

The unbalance current can thus be expressed in terms of (1) the fractional change in unknown resistance, (2) an "effective" detector resistance which depends upon the bridge ratio in use, (3) the sum of all of the bridge resistors, and (4) the applied voltage

Accuracy. The errors in a Wheatstone bridge measurement are caused by (1) the value of unknown resistance and the conditions of measurement, (2) the ability to balance the bridge to the required precision, (3) the available bridge sensitivity, (4) the errors of the comparison resistors, ratios, or both, and (5) an accumulation of small errors resulting from practical circuit and construction problems.

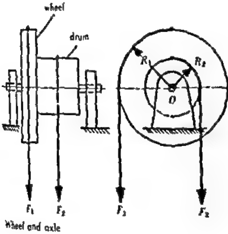
Once the allowable errors of measurement have been determined, the bridge components and the detector should be selected so that the allowable limit of error, the ability to physically adjust the bridge, and the detector sensitivity are in the proportions 1:½:¼. For an allowable error of ±0.1%, the equipment should be adjustable to at least ±0.05%, and the detector should be sufficiently sensitive to detect at least ±0.025% deviation in the measurement.

If the ratio resistors have been adjusted to their individual limits of error, the error in the ratio will probably be larger than the error of each resistor

For this reason, bridge ratio arms are often adjusted for a specified error in ratio, maintaining only a nominal resistance value. [C.T.A.]

Wheel and axle

A wheel and its axle or, more generally, two wheels with different diameters or a wheel and drum, as in a windlass, rigidly connected together so that they



Wheel and axle

rotate as a unit on a common axis. The principle of operation is the same as that of the lever in that, for static equilibrium, the summation of torques about the axis of rotation equals zero. Where flexible members, such as ropes, have been firmly attached to a wheel and drum (as illustrated) and the machine is mounted on frictionless bearings

$$F_1R_1 - F_2R_2 = 0$$

If F_1 is the input and F_2 the output, the mechanical advantage is

$$\mathrm{MA} = \frac{F_2}{F_1} = \frac{R_1}{R_2}$$

If the bearings are not frictionless, a friction torque T_f will oppose the rotation. For the case with F_1 the input, the friction torque will act in the same direction as the load or *output* torque F_2R_2, and the equilibrium equation becomes

$$F_1R_1 - F_2R_2 - T_f = 0$$

and

$$F_2 = \frac{F_1R_1 - T_f}{R_2}$$

Because all members are rigidly connected, all torques act through the same angular displacement, the efficiency is then

$$\eta = \frac{F_2R_2}{F_1R_1} \times 100 = \frac{F_1R_1 - T_f}{F_1R_1} \times 100$$
$$= 1 - \frac{T_f}{F_1R_1} \times 100$$

The main difference between the lever and the wheel and axle is that the wheel and axle permits the forces to operate through a much greater distance. For example, in the illustration the wheel and drum could be allowed to rotate any number of revolutions if the ropes were wrapped the required number of times around each before they were attached.

The wheel and axle is seldom encountered in the exact form shown in the illustration, but the principle of the summation of torques about the axis of rotation equalling zero finds widespread application in machines such as the belt drive, chain drive, friction drive, and gear drive. See SIMPLE MACHINE [R.M.PH.]

Wheel base

The distance in the direction of travel from front to rear wheels of a vehicle, measured between centers of ground contact under each wheel. For a vehicle with two rear axles, the rear measuring point is on the ground midway between rear axles. Tread of a vehicle is the distance perpendicular to the direction of travel between front wheels, or between rear wheels, measured from centers of ground contact.

Wheel bug

A large, gray member, *Arilus cristatus*, of the family Reduviidae, order Hemiptera, characterized by a semicircular crest on the pronotum, which is notched to resemble a cogwheel. These *bugs* are capable of inflicting a very painful bite. Occasionally the victim of such a bite may show serious allergic reactions. The wheel bug is relatively common and is predaceous upon other insects. It is related to the kissing bugs. See HEMIPTERA [J.D.B.]

Whelk

A common name used widely, and apparently indiscriminately in the United States, to identify any of a variety of marine gastropods having *whorled* shells. Originally the name appears to have been applied to certain gastropods of Europe and later to American snails that are sometimes rather distantly related. All whelks have whorled shells and are carnivorous. The name whelk is limited to two families, the Buccinidae and Neptuneidae.

Busycon caricum, known also as *Fulgur carica*, and with a variety of common names including knobbed whelk and knobbed pear conch, is one of the largest and best known. This is the largest gastropod found north of Cape Hatteras, growing in favorable waters to a maximum length of 9 *in*. The pear-shaped shell is large and thick, with six whorls. Its color is yellowish gray, the interior orange red. Those from sheltered waters are not quite as large as specimens from the open beach. It ranges from Cape Cod to Texas. This *powerful* carnivore bores holes in the largest of *oysters*, or *other bivalves*, and sucks their soft parts through the opening, in the manner of the small oyster drills.

The eggs of this whelk are laid in a long parchmentlike ribbon containing as many as 100 egg cases. Such an egg ribbon may be a yard long, and may be laid at any *time* during the warm months in the Atlantic.

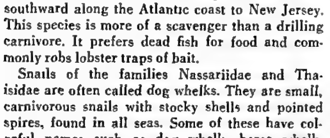
A related large whelk is *Busycon canliculatum* (or *Fulgur caniliculatus*), the channeled whelk, or channeled pear conch It is slightly smaller than *B. caricum*, pale buffy gray in color with a yellow interior. The living pear conch has a yellowish-brown periostracum (outer layer of the shell), bristling with stiff hairs. It is found from Cape Cod to Mexico. It is one of the more active gastropods.

The common edible whelk of Scotland and Ireland is the waved whelk, *Buccinum undatum*. This animal is about 3 in. long, pale reddish to yellowish-brown in color. It is circumpolar and occurs southward along the Atlantic coast to New Jersey. This species is more of a scavenger than a drilling carnivore. It prefers dead fish for food and commonly robs lobster traps of bait.

Snails of the families Nassariidae and Thaisidae are often called dog whelks. They are small, carnivorous snails with stocky shells and pointed spires, found in all seas. Some of these have colorful names such as dog whelk, horse whelk, worn-out basket shell, and tin-roof dog whelk. Several of these prey upon commercially valuable mollusks. *See* GASTROPODA; SNAIL. [J.D.B.]

Whippoorwill

A member, *Caprimulgus vociferus*, of the family Caprimulgidae. This well-known voice of the night sings its name over and over in a fast, crisp whistle. It is not often seen, although it is a characteristic member of the woodland fauna of much of eastern North America, breeding from southern Canada south into northwestern Arkansas and



The whippoorwill, *Caprimulgus vociferus*; length to 10½ in (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

northern Georgia. Like others of the family, it is nocturnal or crepuscular, and feeds on the wing It builds no nest, depositing its eggs instead on the ground *See* CAPRIMULGIFORMES. [J D B]

Whipworm disease

A chronic, wasting diarrhea produced by heavy parasitization of the large intestine by the nematode *Trichuris trichiura* particularly in undernourished children in the tropics. The worms are 1 in. long with the anterior body attenuated to resemble a whip and are firmly attached to the intestine where they irritate the mucosa. Their barrel-shaped eggs pass in the feces. In countries with poor sanitation, the soil becomes contaminated The disease is diagnosed by microscopical examination of the feces.

The eggs embryonate after 3 weeks in soil Upon ingestion, the enclosed larva hatches and arrives directly in the colon where it attains adulthood in 2 months. Light infections, usually acquired by ingesting contaminated foods, are asymptomatic. Heavy ones result from eating of earth, a practice known as geophagy

The disease is treated by oral administration of dithiazanine for a period of 5 days. *See* PARASITOLOGY, MEDICAL; TRICHUROIDEA. [J.F.MA.]

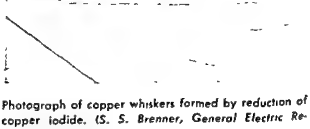
Whiskers, crystal

Single crystals that have grown in a filamentary form. Such filamentary growths have been known for centuries; however, great interest in them developed only after it was discovered at Bell Telephone Laboratories in the 1950s that the strength exhibited by some whiskers in bend tests approaches that expected theoretically for perfect crystals. The strength of whiskers is sometimes as great as 0.06 G, where G is the shear modulus, while the strength of well-annealed single crystals of ordinary experience rarely exceeds 0 0001–0.001 G *See* SINGLE CRYSTAL.

Whiskers have been grown by spontaneous extrusion or out of a supersaturated medium, the medium being supersaturated by physical methods or by chemical reaction. In all instances whiskers apparently form from singularities in the medium

Bell Laboratories scientists discovered that crystal whiskers 1–2 microns in diameter and about 1 mm long are, under certain conditions, extruded from many different metals. It now appears that the motivation for this spontaneous extrusion is relief of internal stress. It has not yet been proven that the spontaneously extruded whiskers, which are strong in bending, are exceptionally strong in tension

Whiskers of a wide variety of substances, for example mercury, graphite, sodium and potassium chlorides, copper, iron, and aluminum oxide, have been grown from supersaturated media Whiskers grown in this way are usually a few microns in diameter and up to several centimeters long (see illustration) Some are exceptionally strong, both in bend tests and in tension tests



Photograph of copper whiskers formed by reduction of copper iodide. (S. S. Brenner, General Electric Research Laboratory)

In addition to exceptional strength, whiskers often have unique electrical, magnetic, or surface properties This behavior can be interpreted to mean that the crystal structure of whiskers is virtually perfect, particularly with respect to line defects. At this writing, however, there is no convincing experimental evidence for this perfection. [D.T.]

Bibliography: S S. Brenner, Growth and properties of "whiskers," *Science*, 128(3324):569–575, 1958, B Chalmers (ed), *Progress in Metal Physics*, vol 6, 1956; R H. Doremus, B. W. Roberts, and D. Turnbull (eds), *Growth and Perfection of Crystals*, 1958

White

One extreme of the series of achromatic colors (see COLOR; COLOR VISION). With respect to the surface of an object, the diffuse, uniform reflectance of all wavelengths characterizes this series from black, through the grays, to white in order of increasing reflectance. Whiteness is a relative term; white paper, paint, and snow reflect some

... object such as a gas or solid at a temperature high enough to emit fairly uniformly light of all visible wavelengths. In the same connotation, a sound is described as white noise if its energy is nearly the same at all audible frequencies. *See* VISION [L.A.R.]

White dwarf star

An intrinsically faint star of very small radius and high density *White* dwarfs occur over a wide range of surface ... as ... s of ... 0 6 ... the Sun, its mean density 3.5×10^5 g/cm³, or about 5 tons/in.³

Under such high pressures, the electrons become degenerate All low electron energy states are occupied, only electrons of exceptionally high energy being able to change their positions and momenta freely The star behaves nearly like a single giant molecule, it is composed largely of helium or heavier elements. *See* DEGENERACY (QUANTUM STATES).

A white dwarf is a final stage of stellar evolution, when all thermonuclear energy sources are extinct. It radiates its internal energy as it cools and becomes fainter About 3% of the stars are now white dwarfs, but in about 5,000,000,000 years all present stars of higher luminosity than the Sun will have burnt out to the white dwarf stage *See* STELLAR EVOLUTION [J L GR.]

Whitefish

Any member of the family Coregonidae, order Isospondyli, a family of silvery fishes with adipose fins and small mouths, without teeth or with very weak ones The whitefishes are circumpolar in distribution. Most species are found only in freshwater although a few spend part of their life in

The whitefish, *Coregonus clupeiformis*, length to 24 in (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

the sea In the United States the most common species, called whitefish, is the Great Lakes whitefish, *Coregonus clupeiformis*. Prior to the depredations by the marine lamprey, this was the most important food fish produced in the Great Lakes. It still supports a substantial fishery in Canada. This species ranges northward into the Arctic Circle and is a prized food fish, fresh or smoked. Its roe is sold as caviar. This whitefish averages two or three pounds in weight. *See* CLUPEIFORMES [J.D B.]

Whooping cough

An acute lung infection characterized by paroxysmal coughing It is also known as pertussis and is caused by the bacterium *Bordetella pertussis* which, until 1957, was called *Haemophilus pertussis* The microbiological diagnosis of pertussis is made by growing the organisms on Bordet Gengou agar containing 20–30% blood Material may be coughed directly upon the agar A preferable procedure obtains material by swabbing the posterior nasal passages through the nose with a cotton swab cemented to a flexible wire.

Whooping cough has an incubation period of 7–14 days and is highly contagious. Of all infectious diseases, it is the most frequent cause of death in the first 18 months of life, but later the death rate is low. Treatment is not very effective, but three doses of pertussis vaccine at monthly intervals beginning soon after birth will protect most babies, and booster or reinforcing doses at 1 and 4 years will usually protect completely *See* BIOLOGICALS; BRUCELLACEAE, HEMOPHILIC BACTERIA [W F.V.]

Wiedemann-Franz law

An empirical law of physics which states that the ratio of the thermal conductivity of a metal to its electrical conductivity is a constant times the absolute temperature:

$$K_e = L_0 \sigma T$$

Here K_e is the thermal conductivity due to the conduction electrons, σ is the electrical conductivity, T the absolute temperature, and L_0 is known as the Lorenz number For the case of a degenerate electron gas, the value of L_0 is given by

$$L_0 = (\pi^2/3)(k/e)^2$$

where k is the Boltzmann constant and where e is

the electronic charge. The Wiedemann-Franz law provides an important check on theories of electrical and thermal conductivity. *See* CONDUCTION (HEAT); ELECTRICAL CONDUCTIVITY OF METALS; FREE-ELECTRON THEORY OF METALS. [F.J.B.]

Wigner-Seitz method

The Wigner-Seitz, or cellular, method is a technique for calculating the energy levels of electrons in solids, based on the expansion of the wave function of an electron in an atomic cell in a series of products of radial functions and spherical harmonics. It was developed by E. Wigner and F. Seitz in 1933.

Consider, for simplicity, a monatomic solid. The atomic cell is constructed in the following way: a line is drawn connecting one atom and all its neighbors The planes which are the perpendicular bisectors of these lines are then constructed These planes bound a solid figure which encloses one atom.

The problem consists of solving Schrödinger's equation, or, more precisely, the Hartree-Fock equation, in this cell subject to the appropriate boundary conditions on its surface The basic idea is to expand the wave function in terms of functions which satisfy the Hartree-Fock equation exactly, but not the boundary conditions, and then to adjust the coefficients in the expansion so that the boundary conditions are satisfied. The potential energy of an electron is assumed to be spherically symmetric within such a cell.

A cellular-method calculation of energy levels in solids involves a very considerable amount of numerical computation in regard to integration of the radial equation and in the application of the boundary conditions. Use of an electronic computer is required. The cellular method, however, is not the only practical procedure for the study of energy bands. There are other approaches, notably the orthogonalized plane-wave (OPW) method, which can give reliable results with no more labor. The OPW method involves expansion of the solid-state wave function in terms of functions which satisfy the exact boundary conditions but solve the differential equation only approximately. The coefficients are adjusted to give a solution of the Schrödinger equation.

Application of the Wigner-Seitz method has been made to many simple substances, including among others lithium, sodium, copper, silicon, iron, lithium fluoride, zinc sulfide, and barium oxide A particularly interesting application of the cellular method, the quantum-defect method, has been developed by H. Brooks and others. In this approach, which has been successfully applied to all the alkali metals, information obtained from the spectra of free atoms is used directly in the construction of the solid-state wave function. *See* BAND THEORY OF SOLIDS; QUANTUM THEORY, NONRELATIVISTIC. [J.C.]

Bibliography: F. Seitz and D. Turnbull (eds.), *Solid State Physics*, vol. 1, 1955, vol. 7, 1958.

Wildlife conservation

Wildlife, in a restricted sense, refers to undomest cated, warm-blooded vertebrates, or wild mam mals and birds. Two groups are important, gam and nongame species. Game species can be sul divided into big-game mammals, small-game mam mals, furbearers, upland game birds, marsh bird and waterfowl (ducks and geese). Nongame form include those mammals and birds that lack spo value

Wildlife serves man in many ways. The searc for valuable peltries, such as beaver, motivate much of the exploration of the New Worl (Fig. 1). At one time the beaver ranged over nin tenths of North America. Today this species down to a few hundred thousand animals, largel extirpated from much of its former range. Recer beaver restoration attempts, however, are retur ing small colonies of beaver to limited areas.

The bison provided food, clothing, and she ter for many Indians (Fig. 2). Countless bir were once exploited for plumage and eggs, whi others such as passenger pigeons, quail, shorebird ducks, and geese were hunted for the market.

Today in a complex, high-pressure society, hun ing provides exercise and relaxation for million One survey showed that 12,000,000 Americans a nually spend $937,000,000 on hunting An annu $5,000,000,000 fur business and the meat value c game also add materially to the nation's econom Of great importance also are the multiple soci benefits associated with wildlife That wild cre tures have a great capacity for inspiring creati reaction to their beauty is evident in works of ar music, and poetry

Wildlife is important, too, for the role it plays i the ecological balance of nature (*see* ECOLOGY For example, when certain predatory mammal (Fig 3) and birds are killed or driven from a area, there soon occurs an unusual abundance c harmful animals formerly preyed upon, such as r dents. Some mammals and birds distribute an plant seeds and nuts. Many birds consume wee

Fig. 1. Beaver, an important fur animal almost ex terminated by unregulated trapping

Fig 2. American bison, a victim of ruthless slaughter during the opening of the western plains to settlement (USDA)

seeds and some eat large quantities of injurious insects.

Conservation problems. In the early days North America was a paradise for many forms of wildlife. American bison, antelope, deer, elk, moose, mountain sheep, caribou, musk oxen, and the large carnivores, such as bears, mountain lions, and wolves, ...ted over extensive areas. Small game flourished ...terfowl darkened the skies (Fig. 4). Today, ...h few exceptions, notably the whitetailed deer ... antelope, big game populations have drasti...y declined. Many birds have been reduced in ...nbers. Some have become extinct.

...f the many contributing causes of depletion, ...two most important are the removal of proper ...ironment (Fig 5) by the destruction of forests, ...tlands, and grasslands by man and his imple...nts (see FOREST CONSERVATION; RANGE LAND ...SERVATION) and the direct decimation of the ...cies by overshooting and overtrapping. The ...ge buffalo herds disappeared because of over...ting and range destruction, the beaver de...ned because of overtrapping, and the grizzly ...ar is nearly extinct because of overhunting ...ecies like the heath hen and ivory-billed wood...cker nearly *vanished* because man seriously dis...rbed their environment.

Management and conservation measures. ...merica began to awaken to the plight of its wild...fe late in the nineteenth century. Gradually pro-

Fig. 3. The ... prevent ... other predatory animals help ... undesirable species

Fig 4 Ducks and geese were once abundant throughout most of the United States. (After J. Von Huizen, U.S. Fish and Wildlife Service)

Fig 5 Destruction of wildlife habitat contributes greatly to reduction in wildlife numbers

tection measures were undertaken, followed shortly by restoration attempts. Following World War I, many thought that the way to game abundance was through ingenious methods of artificial propagation, and game farms sprang up all around the country (Fig 6). However, the fallacy of this hand-rearing of game soon became apparent. Experience and research proved that the stocking of pen-raised game in areas where native stock was present, but not doing well, was a wasteful practice. A few states still maintain game farms because of demand from a public unaware of the facts. Today *the game farm is a tool of game management to be* used only when justified, such as in the raising of brood stock for areas where no parent stock exists.

Present day wildlife management looks upon game as a crop with an annual surplus to be harvested. Game laws and regulations seek to restrict reduction in numbers to the safe allowable annual surplus. Game agencies today rely on sportsmen's views and on inventories by trained biologists in determining seasons and bag limits.

Predator control, disease control, and refuges were also once thought to be the answer to game abundance but now are considered only tools of the game manager. Predator control has been relegated to a practice useful only in extreme or local situa-

Fig 6 Game farm (Virginia Commission of Game and Inland Fisheries)

tions. Disease control in wildlife poses many difficulties, the best insurance being abundant food, cover, and range. Refuges serve mostly as an adjunct to good land management.

Today wildlife conservation is largely being directed at (1) legal measures to restrict the annual kill by imposition of licenses, open and closed seasons and hours, and daily and seasonal bag limits, including the taking of certain sexes, and (2) the manipulation of environment to encourage the growth of certain animal populations. The growing of certain food and cover plants required by game species through multiple-purpose land use is also an essential part of wildlife management (Fig. 7).

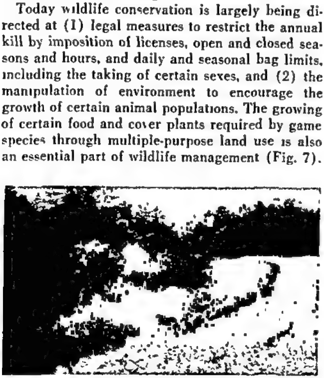


Fig 7 Food and cover plants are essential to wildlife restoration. (USDA)

Action programs. In all parts of America, Federal, state, and provincial governments and some private organizations are engaged in wildlife conservation work Protecting migratory wildlife is the main responsibility of the U.S Fish and Wildlife Service, which handles Federal law enforcement, game inventories, research, predator control, and wildlife management on some 200 refuges (Fig. 8) An outstanding event was the passage of the Pittman-Robertson Act of 1937, which levied an 11% Federal tax on guns and ammunition. The revenue from this tax is returned to the states on the basis of 75% Federal funds to 25% state matching funds and is being used for approved wildlife projects in research and management.

Other Federal agencies active in wildlife conservation work are the Forest Service (on national forests totaling 176,000,000 acres), the Soil Conservation Service, and the National Park Service In the provinces and states, wildlife responsibility is assumed chiefly by conservation departments or wildlife commissions supported by funds from hunting licenses These departments or commissions administer game laws, carry on wildlife research and education, and work with other agencies and groups in promoting land-management practices favoring wildlife.

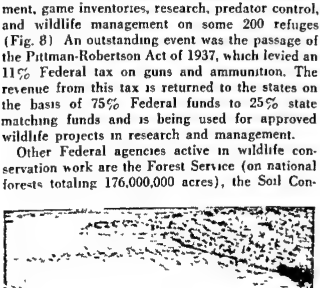


Fig. 8 Migratory wildlife must be protected to prevent extermination

Private conservation groups, such as the Izaak Walton League of America, National Wildlife Federation, Ducks Unlimited, and the National Audubon Society, also aid the wildlife conservation effort. Work of the Wildlife Management Institute in supporting research and education and in promoting the annual North American Wildlife Conferences is likewise significant, as is the work of such organizations as the Wildlife Society, the International Association of Fish and Game Conservation Commissioners, the several regional associations of fish and game officials, and the flyway councils. [J.J.S.]

Bibliography: *See* CONSERVATION OF RESOURCES

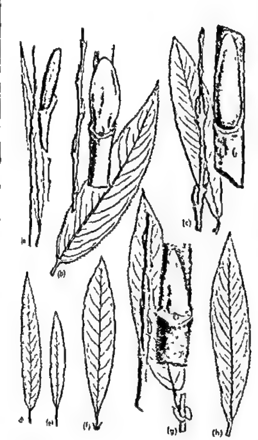
Willemite

A nesosilicate mineral, composition Zn_2SiO_4, crystallizing in the hexagonal system It is usually massive or granular with a vitreous luster; crystals are rare. The mineral has a basal cleavage and may be variously colored, most commonly green, red, or brown. Hardness is 5½ on Mohs scale; specific gravity is 3 9–4 2. Willemite forms a valuable ore of zinc at Franklin, New Jersey. At this famous zinc deposit willemite fluoresces a yellow-green and is found in crystalline limestone associated with franklinite, zincite, and many other rarer minerals. With the exception of the occurrence in New Jersey, willemite is a rare mineral. *See* SILICATE MINERALS [C.S.HU.]

Willow

A deciduous tree and shrub of the genus *Salix*, order Salicales, common along streams and in wet places in the United States, Europe, and China (*see* SALICALES) The twigs are often yellow-green and bear alternate leaves which are characteristically long, narrow, and pointed, usually with fine teeth along the margins. *See* LEAF (BOTANY). Flowers occur in catkins, the staminate consist of 1–12 (usually 2) stamens, with 1 or 2 nectar glands, the pistillate have a single ovary of 2 carpels, with 1–4 small glands. *See* FLOWER (BOTANY). The fruit contains several silky seeds *See* FRUIT (BOTANY); SEED (BOTANY). The species are wind or insect-pollinated and hybridize freely, which often makes identification difficult.

The annual average production of willow lumber from 1933–1942 was approximately 18,000,000 board ft. The wood is used for fuel and in making charcoal, excelsior, ball bats, boxes, crates, boats, waterwheels, and wicker furniture However, the wood often warps badly in curing unless great care is taken, and it is usually too brittle to be used for heavy construction The tough, pliable shoots of



(a) Babylon weeping willow, *Salix babylonica* (b) Crack willow, *Salix fragilis* (c) Purple-osier willow, *Salix purpurea* (d) Silky willow, *Salix sericea*. (e) Sage willow, hoary willow, *Salix candida* (f) Prairie willow, *Salix humilis*. (g) Heartleaf willow, *Salix cordata* (h) Yellow-stem white willow, *Salix alba* var *vitellina* (Reprinted from A H Graves, *Illustrated Guide to Trees and Shrubs*, rev ed, Harper, 1956)

many species are used to make baskets, the bark of other species is used for tanning. Willows are of great value in checking soil erosion. A few species, such as the white willow and weeping willow, are ornamental shade trees. See TREE. [J.F F]

Wind

The motion of air relative to the earth's surface. The term usually refers to *horizontal* air motion, as distinguished from vertical motion, and to air motion averaged over a chosen period of 1–3 min. Micrometeorological circulations (air motion over periods of the order of a few seconds) and others small enough in extent to be *obscured* by this averaging are thereby eliminated. The choice of the 1- to 3-min interval has proven suitable for the study of (1) the hour-to-hour and day-to-day changes in the atmospheric circulation pattern, and (2) the larger scale aspects of the atmospheric general circulation.

The direct effects of wind near the surface of the earth are manifested by soil erosion, the character of vegetation, damage to structures, and the production of waves on water surfaces. At higher levels wind directly affects aircraft, missile and rocket operations, and dispersion of industrial pollutants, radioactive products of nuclear explosions, dust, volcanic debris, and other material. Directly or indirectly, wind is responsible for the production and transport of clouds and precipitation and for the transport of cold and warm air masses from one region to another. *See* CONVECTION (HEAT).

Wind direction and speed. Since wind is a vector quantity, both of these aspects must be specified if the wind is to be described completely. Wind direction is designated as the geographical direction from which the air is moving. Direction is specified either in terms of the conventional letter abbreviations of cardinal compass directions, or in terms of the number of angular degrees of clockwise departure from north (azimuth—north being equivalent to 0° or 360°, east to 90°, south to 180°, and west to 270°).

Wind speed (often referred to loosely as wind velocity) is most frequently expressed in units of nautical miles per hour (knots) or in units of meters per second. Average wind speed near the surface of the earth is between 10 and 15 knots, but depends strongly upon geographical location and season. Much higher wind speeds are produced temporarily by various types of storms. Wind speeds at higher levels are generally greater than those near the surface, particularly in middle latitudes. *See* JET STREAM, STORM.

Measurement of wind. Direct methods near the surface of the earth and indirect methods at higher levels are used to measure wind. The surface wind direction and speed are obtained from the indications of a wind vane and anemometer (usually of the rotating-cup or pressure-tube type) mounted a few tens of feet above the surface and well away from obstructions to the flow of air. Aloft measurements of wind are commonly obtained by tracking the motion of a balloon which is presumed to be moving freely with the air stream. The tracking is accomplished from the ground either by visual methods (through use of one or more theodolites), or by radar or other radio direction finding techniques.

The Beaufort wind scale may be used to estimate the speed of the surface wind by observation of the effect of the wind upon trees, flags, dust, smoke plumes, or water bodies. Since 1949 its application has been largely restricted to observations by amateurs and to observations aboard ships not equipped with anemometers. Some of the descriptive terminology, however, remains in widespread [illegible]

ar [illegible]

5 [illegible]

50 knots, storm a speed between 50 and 65 knots, and hurricane a speed equal to or in excess of 65 knots. *See* METEOROLOGICAL INSTRUMENTATION, THEODOLITE.

Cyclonic and anticyclonic circulation. Each is a portion of the pattern of air flow within which the streamlines (which indicate the pattern of wind direction at any instant) are curved so as to indicate

opposite rotation of air about some central point of the cyclone or anticyclone. Thus, in a cyclonic circulation, the streamlines indicate counterclockwise (clockwise for anticyclonic) rotation of air about a central point on the Northern Hemisphere or clockwise (counterclockwise for anticyclonic) rotation about a point on the Southern Hemisphere. When the streamlines close completely about the central point, the pattern is denoted respectively a cyclone or an anticyclone. Since the gradient wind represents a good approximation to the actual wind, the center of a cyclone tends strongly to be a point of minimum atmospheric pressure on a horizontal surface. Thus the terms cyclone, low-pressure area, or "low" are often used to denote essentially the same phenomenon. In accord with the requirements of the gradient wind relationship, the center of an anticyclone tends to coincide with a point of maximum pressure on a horizontal surface, and the terms anticyclone, high-pressure area, or "high" are often used interchangeably.

Cyclones and anticyclones are numerous in the lower troposphere at all latitudes. At higher levels, the occurrence of cyclones and anticyclones tends to be restricted to subpolar and subtropical latitudes respectively. In middle latitudes the flow aloft is mainly westerly, but the streamlines exhibit wavelike oscillations connecting adjacent regions of anticyclonic circulation (ridges) and of cyclonic circulation (troughs).

Although the atmosphere is never in a completely undisturbed state, it is customary to refer to cyclonic and anticyclonic circulations specifically as atmospheric disturbances. Cyclones, anticyclones, ridges, and troughs are intimately associated with the production and transport of clouds and precipitation, and hence convey a connotation of disturbed meteorological conditions.

A more rigorous definition of circulation is often employed, in which the circulation over an arbitrary area bounded by the closed curve S is given by

$$C = \oint v_t \, dS$$

in which the integration is taken completely around the boundary of the area, v refers to the wind at a point on the boundary, the subscript t denotes the component of this wind parallel to the boundary and dS is a line element of the boundary. The component v_t is considered positive or negative accordingly as it represents cyclonic or anticyclonic circulation along the boundary S. In this context, the circulation may be positive (cyclonic) or negative (anticyclonic) even when the streamlines within the area are straight, since the distribution of wind speed affects the value of C. *See* AIR WAVES, UPPER SYNOPTIC; ATMOSPHERE; ATMOSPHERIC HIGH; ATMOSPHERIC LOW; CLOUD; GEOSTROPHIC WIND; GRADIENT WIND; PRECIPITATION (METEOROLOGY); STORM.

Convergent or divergent patterns. These are said to occur in areas in which the (horizontal) wind flow and distribution of air density is such as to produce a net accumulation or depletion, respectively, of mass of air. Rigorously, the mean horizontal mass divergence over an arbitrary area A bounded by the closed curve S is given by

$$D = \frac{1}{A} \oint \rho v_n \, dS$$

in which the integration is taken completely around the boundary of the area, ρ is the density of air, v refers to the wind at a point on the boundary, the subscript n denotes the component of this wind perpendicular to the boundary, and dS is an element of the boundary. The component v_n is taken positive when it is directed outward across the boundary and negative when it is directed inward. Convergence is thus synonymous with negative divergence. If spatial variations of density are neglected, the analogous concept of velocity divergence and convergence applies.

The horizontal mass divergence or convergence is intimately related to the vertical component of motion. For example, since local temporal rates of change of air density are relatively small, there must be a net vertical export of mass from a volume in which horizontal mass convergence is taking place. Only thus can the total mass of air within the volume remain approximately constant. In particular, if the lower surface of this volume coincides with a level ground surface, upward motion must occur across the upper surface of this volume. Similarly, there must be downward motion immediately above such a region of horizontal mass divergence.

The horizontal mass divergence or convergence is closely related to the circulation. In a convergent wind pattern the circulation of the air tends to become more cyclonic; in a divergent wind pattern the circulation of the air tends to become more anticyclonic.

Regions which lie in the path of an approaching cyclone are characterized by a convergent wind pattern in the lower troposphere and by upward vertical motion throughout most of the troposphere. Since the upward motion tends to produce condensation of water vapor in the rising air current, abundant cloudiness and precipitation typically occur in this region. Conversely, the area in advance of an anticyclone is characterized by a divergent wind pattern in the lower troposphere and by downward vertical motion throughout most of the troposphere. In such a region, clouds and precipitation tend to be scarce or entirely lacking.

A convergent surface wind field is typical of fronts. As the warm and cold currents impinge at the front, the warm air tends to rise over the cold air producing the typical frontal band of cloudiness and precipitation. *See* FRONT.

Zonal surface winds. Such patterns result from a longitudinal averaging of the surface circulation. This averaging typically reveals a zone of weak variable winds near the Equator (the doldrums) flanked by northeasterly trade winds in the Northern Hemisphere and southeasterly trade winds on the Southern Hemisphere, extending poleward in each instance to about latitude 30°. The doldrum

belt, particularly at places and times at which it is so narrow that the trade winds from the two hemispheres impinge upon it quite sharply, is designated the intertropical convergence zone, or ITC. The resulting convergent wind field is associated with abundant cloudiness and locally heavy rainfall. A westerly average of zonal surface winds prevails poleward of the trade wind belts and dominates the middle latitudes of both hemispheres. The westerlies are separated from the trade winds by the subtropical high pressure belt, which occurs between latitudes 30° and 35° (the horse latitudes), and are bounded on the poleward side in each hemisphere between latitudes 55° and 60° by the subpolar trough of low pressure. Numerous cyclones and anticyclones progress eastward in the zone of prevailing westerlies, producing the abrupt day-to-day changes of wind, temperature, and weather which typify these regions. Poleward of the subpolar low pressure troughs, polar easterlies are observed.

The position and intensity of the zonal surface wind systems vary systematically from season to season and irregularly from week to week. In general the systems are most intense and are displaced toward the Equator in a given hemisphere during winter. In this season the subtropical easterlies and prevailing westerlies attain mean speeds of about 15 knots, while the polar easterlies are somewhat weaker. In summer the systems are displaced toward the pole by 5° or 10° of latitude and weaken to about one-half their winter strength.

When the pattern of wind circulation is averaged with respect to time instead of longitude, striking differences between the Northern and Southern Hemispheres are found. On the Southern Hemisphere, variations from longitude to longitude are relatively small and the averaged pattern is described quite well in terms of the zonal surface wind belts. On the Northern Hemisphere there are large differences from longitude to longitude. In winter, for example, the subpolar trough is mainly manifested in two prominent low centers, the Icelandic low and the Aleutian low. The subtropical ridge line is drawn northward in effect over the continents and is seen as a powerful and extensive high-pressure area over Asia and as a relatively weak area of high pressure over North America. In summer the Aleutian and Icelandic lows are weak or entirely absent, while extensive areas of low pressure over the southern portions of Asia and western North America interrupt the subtropical high pressure belt. See MONSOON.

Upper air circulation. Longitudinal averaging indicates a predominance of westerly winds. These westerlies typically increase with elevation and culminate in the average jet stream, which is found in lower middle latitudes near the tropopause at elevations between 35,000 and 40,000 ft. The subtropical ridge line aloft is found equatorward of its surface counterpart and easterlies occur at upper levels over the equatorward portions of the trade wind belts. In high latitudes, weak westerlies aloft are found over the surface polar easterlies. Seasonal and irregular fluctuations of the circulation aloft are similar to those which characterize the surface winds. See AIR WAVES, UPPER SYNOPTIC; ATMOSPHERE; JET STREAM.

Minor terrestrial winds. In this category are circulations of relatively small scale, attributable indirectly to the character of the earth's surface. One example, the land and sea breeze, is a circulation driven by pronounced heating or cooling of a given area in comparison with little heating or cooling in a horizontally adjacent area. During the day, air rises over the strongly heated land and is replaced by a horizontal breeze from the relatively cool sea. At night, air sinks over the cool land and spreads out over the now relatively warm sea.

Another example is formed by the mountain and valley winds. These result from cooling and heating, respectively, of the mountain slopes relative to the horizontally adjacent free air above the valley floor. During the day, air flows up from the valley along the strongly heated mountain slopes, but at night, air flows down the relatively cold mountain slopes toward the valley bottom. A similar type of descending current of cooled air is often observed along the sloping surface of a glacier. This nighttime air drainage, under proper topographical cir- [illegible]

[illegible] cations by local topography of a circulation of large scale. They are often capricious and violent in nature and are sometimes characterized by extremely low relative humidity. Examples are the mistral which blows down the Rhone Valley in the south of France, the bora which blows down the gorges leading to the coast of the Adriatic Sea, the foehn winds which blow down the Alpine valleys, [illegible]

[illegible] given in some instances to currents of somewhat larger scale which are less directly related to topography. Examples of this type of wind are the norther, which represents the rapid flow of cold air from Canada down the plains east of the Rockies and along the east coast of Mexico into Central America; the nor'easter of New England, which is part of the wind circulation about intense cyclones [illegible]

[illegible] North Africa and sometimes crosses the Mediterranean Sea. See CHINOOK; SIROCCO. [F S.]

Wind measurement

The determination of three parameters: the size of an air sample, its speed, and its direction of motion.

Size of sample. The size of the air sample is highly dependent on how the measurement is made. When the wind measurement is taken with small,

opposite rotation of air about some central point of the cyclone or anticyclone. Thus, in a cyclonic circulation, the streamlines indicate counterclockwise (clockwise for anticyclonic) rotation of air about a central point on the Northern Hemisphere or clockwise (counterclockwise for anticyclonic) rotation about a point on the Southern Hemisphere. When the streamlines close completely about the central point, the pattern is denoted respectively a cyclone or an anticyclone. Since the gradient wind represents a good approximation to the actual wind, the center of a cyclone tends strongly to be a point of minimum atmospheric pressure on a horizontal surface. Thus the terms cyclone, low-pressure area, or "low" are often used to denote essentially the same phenomenon. In accord with the requirements of the gradient wind relationship, the center of an anticyclone tends to coincide with a point of maximum pressure on a horizontal surface, and the terms anticyclone, high-pressure area, or "high" are often used interchangeably.

Cyclones and anticyclones are numerous in the lower troposphere at all latitudes. At higher levels, the occurrence of cyclones and anticyclones tends to be restricted to subpolar and subtropical latitudes, respectively. In middle latitudes the flow aloft is mainly westerly, but the streamlines exhibit wavelike oscillations connecting adjacent regions of anticyclonic circulation (ridges) and of cyclonic circulation (troughs).

Although the atmosphere is never in a completely undisturbed state, it is customary to refer to cyclonic and anticyclonic circulations specifically as atmospheric disturbances. Cyclones, anticyclones, ridges, and troughs are intimately associated with the production and transport of clouds and precipitation, and hence convey a connotation of disturbed meteorological conditions.

A more rigorous definition of circulation is often employed, in which the circulation over an arbitrary area bounded by the closed curve S is given by

$$C = \oint v_t \, dS$$

in which the integration is taken completely around the boundary of the area, v refers to the wind at a point on the boundary, the subscript t denotes the component of this wind parallel to the boundary and dS is a line element of the boundary. The component v_t is considered positive or negative accordingly as it represents cyclonic or anticyclonic circulation along the boundary S. In this context the circulation may be positive (cyclonic) or negative (anticyclonic) even when the streamlines within the area are straight, since the distribution of wind speed affects the value of C. See AIR WAVES, UPPER SYNOPTIC; ATMOSPHERE; ATMOSPHERIC HIGH; ATMOSPHERIC LOW; CLOUD; GEOSTROPHIC WIND; GRADIENT WIND; PRECIPITATION (METEOROLOGY); STORM.

Convergent or divergent patterns. These are said to occur in areas in which the (horizontal) wind flow and distribution of air density is such as to produce a net accumulation or depletion, respectively, of mass of air. Rigorously, the mean horizontal mass divergence over an arbitrary area A bounded by the closed curve S is given by

$$D = \frac{1}{A} \oint \rho v_n \, dS$$

in which the integration is taken completely around the boundary of the area, ρ is the density of air, v refers to the wind at a point on the boundary, the subscript n denotes the component of this wind perpendicular to the boundary, and dS is an element of the boundary. The component v_n is taken positive when it is directed outward across the boundary and negative when it is directed inward. Convergence is thus synonymous with negative divergence. If spatial variations of density are neglected, the analogous concept of velocity divergence and convergence applies.

The horizontal mass divergence or convergence is intimately related to the vertical component of motion. For example, since local temporal rates of change of air density are relatively small, there must be a net vertical export of mass from a volume in which horizontal mass convergence is taking place. Only thus can the total mass of air within the volume remain approximately constant. In particular, if the lower surface of this volume coincides with a level ground surface, upward motion must occur across the upper surface of this volume. Similarly, there must be downward motion immediately above such a region of horizontal mass divergence.

The horizontal mass divergence or convergence is closely related to the circulation. In a convergent wind pattern the circulation of the air tends to become more cyclonic; in a divergent wind pattern the circulation of the air tends to become more anticyclonic.

Regions which lie in the path of an approaching cyclone are characterized by a convergent wind pattern in the lower troposphere and by upward vertical motion throughout most of the troposphere. Since the upward motion tends to produce condensation of water vapor in the rising air current, abundant cloudiness and precipitation typically occur in this region. Conversely, the area in advance of an anticyclone is characterized by a divergent wind pattern in the lower troposphere and by downward vertical motion throughout most of the troposphere. In such a region, clouds and precipitation tend to be scarce or entirely lacking.

A convergent surface wind field is typical of fronts. As the warm and cold currents impinge at the front, the warm air tends to rise over the cold air producing the typical frontal band of cloudiness and precipitation. See FRONT.

Zonal surface winds. Such patterns result from a longitudinal averaging of the surface circulation. This averaging typically reveals a zone of weak variable winds near the Equator (the doldrums) flanked by northeasterly trade winds in the Northern Hemisphere and southeasterly trade winds on the Southern Hemisphere, extending poleward in each instance to about latitude 30°. The doldrum

belt, particularly at places and times at which it is so narrow that the trade winds from the two hemispheres impinge upon it quite sharply, is designated the intertropical convergence zone, or ITC. The resulting convergent wind field is associated with abundant cloudiness and locally heavy rainfall. A westerly average of zonal surface winds prevails poleward of the trade wind belts and dominates the middle latitudes of both hemispheres. The westerlies are separated from the trade winds by the subtropical high pressure belt, which occurs between latitudes 30° and 35° (the horse latitudes), and are bounded on the poleward side in each hemisphere between latitudes 55° and 60° by the subpolar trough of low pressure. Numerous cyclones and anticyclones progress eastward in the zone of prevailing westerlies, producing the abrupt day-to-day changes of wind, temperature, and weather which typify these regions. Poleward of the subpolar low pressure troughs, polar easterlies are observed.

The position and intensity of the zonal surface wind systems vary systematically from season to season and irregularly from week to week. In general the systems are most intense and are displaced toward the Equator in a given hemisphere during winter. In this season the subtropical easterlies and prevailing westerlies attain mean speeds of about 15 knots, while the polar easterlies are somewhat weaker. In summer the systems are displaced toward the pole by 5° or 10° of latitude and weaken [illegible]

[illegible] differences between the Northern and Southern Hemispheres are found. On the Southern Hemisphere, variations from longitude to longitude are relatively small and the averaged pattern is described quite well in terms of the zonal surface wind belts. On the Northern Hemisphere there are large differences from longitude to longitude. In winter, for example, the subpolar trough is mainly manifested in two prominent low centers, the Icelandic low and the Aleutian low. The subtropical ridge line is drawn northward in effect over the continents and is seen as a powerful and extensive high pressure area over Asia and as a relatively weak area of high pressure over North America. In summer the Aleutian and Icelandic lows are weak or entirely absent, while extensive areas of low pressure over the southern portions of Asia and western North America interrupt the subtropical high pressure belt. See MONSOON.

Upper air circulation. Longitudinal averaging indicates a predominance of westerly winds. These westerlies typically increase with elevation and culminate in the average jet stream, which is found in lower middle latitudes near the tropopause at elevations between 35,000 and 40,000 ft. The subtropical ridge line aloft is found equatorward of its surface counterpart and easterlies occur at upper levels over the equatorward portions of the trade wind belts. In high latitudes, weak westerlies aloft are found over the surface polar easterlies. Seasonal and irregular fluctuations of the circulation aloft are similar to those which characterize the surface winds. *See* AIR WAVES, UPPER SYNOPTIC; ATMOSPHERE; JET STREAM.

Minor terrestrial winds. In this category are circulations of relatively small scale, attributable indirectly to the character of the earth's surface. One example, the land and sea breeze, is a circulation driven by pronounced heating or cooling of a given area in comparison with little heating or cooling in a horizontally adjacent area. During the day, air rises over the strongly heated land and is replaced by a horizontal breeze from the relatively cool sea. At night, air sinks over the cool land and spreads out over the now relatively warm sea.

Another example is formed by the mountain and valley winds. These result from cooling and heating, respectively, of the mountain slopes relative to the horizontally adjacent free air above the valley floor. During the day, air flows up from the valley along the strongly heated mountain slopes, but at night, air flows down the relatively cold mountain slopes toward the valley bottom. A similar type of descending current of cooled air is often observed along the sloping surface of a glacier. This nighttime air drainage, under proper topographical cir- [illegible]

[illegible] cations by local topography of a circulation of large scale. They are often capricious and violent in nature and are sometimes characterized by extremely low relative humidity. Examples are the mistral which blows down the Rhone Valley in the south of France, the bora which blows down the gorges leading to the coast of the Adriatic Sea, the foehn winds which blow down the Alpine valleys, the williwaws which are characteristic of the fiords of the Alaskan coast and the Aleutian Islands, and the chinook which is observed on the eastern slopes of the Rocky Mountains. Local names are also given in some instances to currents of somewhat larger scale which are less directly related to topography. Examples of this type of wind are the norther, which represents the rapid flow of cold air from Canada down the plains east of the Rockies and along the east coast of Mexico into Central America; the nor'easter of New England, which is part of the wind circulation about intense cyclones centered offshore along the middle Atlantic coastal states; and the sirocco, a southerly wind current from the Sahara which is common on the coast of North Africa and sometimes crosses the Mediterranean Sea. See CHINOOK, SIROCCO. [F.S.]

Wind measurement

The determination of three parameters, the size of an air sample, its speed, and its direction of motion.

Size of sample. The size of the air sample is highly dependent on how the measurement is made. When the wind measurement is taken with small,

sensitive, and rapid-response instruments, or by the drift of small suspended particles, the air sample can have a scale of millimeters or less. If the wind measurement is made from the pressure gradient (*see* GEOSTROPHIC WIND; GRADIENT WIND), as measured on a weather map, the scale can be hundreds of kilometers. Wind measurements also depend on the system used. When the wind is measured at a point fixed with respect to the earth's surface with air moving by, the measurement is called Eulerian. In this, one is continuously measuring different air samples. Continuing measure of the speed and direction of the same air sample is called a Lagrangian measurement. This is obtained by measuring the drift of balloons or smoke. The two systems only give the same value when the wind is perfectly steady. Ordinarily when a wind speed or direction is given, some sort of averaging, usually a time average, is implied. *See* TURBULENT FLOW.

Wind direction. This is designated as the direction from which the wind is blowing, given in terms of 8, 16, or 32 points of the compass and in degrees or tens of degrees from north, measured clockwise. A calm wind is reported at 00; a north wind is 36 (representing 360°) A wind vane is used to measure the direction of the surface wind. Basically, this is a grossly unsymmetrical body mounted near its center of gravity and free to rotate about a vertical axis. This simple and old meteorological instrument now usually includes an electrical device for remote indication or recording of the wind direction A wind vane free to rotate about a vertical and a horizontal axis (called a bivane) also indicates the vertical wind component

Wind speed. In terms of its force on common objects such as leaves, smoke, or waves, wind speed can be estimated by the Beaufort scale Meteorologists often measure the wind speed in terms of the cause of the wind, that is, the atmospheric pressure gradient, which is inversely proportional to the spacing of the isobars on a weather map. For information concerning instruments which measure surface winds, *see* ANEMOMETER, for further information on this and on measurement of upper-air wind speed, *see* METEOROLOGICAL INSTRUMENTATION; *see also* TELEMETERING. [V E.S.]

Bibliography: W E. K. Middleton, *Meteorological Instruments*, 3d ed , 1953.

Wind power

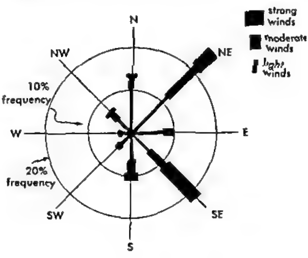
In the late 1950s over 300,000 pumping windmills and nearly 100,000 wind-driven electric plants were in operation in the United States, but the energy generated by these little plants totaled only about 400,000,000 hp-hr/year. This is less than a third of the energy that is used to operate electric fans in homes and industry.

It has been estimated that the total energy of the wind amounted to around 15% of the world's energy requirement (5×10^{13} hp-hr) in 1957, but wind energy is kinetic, and if an attempt were made to capture all of it, the wind would disappear. The proportion of the potential that can be recovered exceedingly small because of a number of practic considerations. Wind is highly variable, with pe ods of calm and periods of excess Windmil rugged enough to withstand high winds can pr duce no net power from moderate winds. Win much under 20 miles per hour (mph) has not bee considered practical, and wind much over 30 mp is unmanageable Wind velocity is 275% great 500 ft above the surface of the earth than at groun level, but the presence of very high towers in th air age is not desirable. Because of variations i wind velocity, windmills are not capable of delive ing firm power.

In spite of such handicaps, much serious though has been given to possibilities of commercial win power development. For example, it has been est mated that in Scotland several hundred suitabl sites might be found for 2000-kw wind generato which might give a total of around 4,000,000,00 kw-hr annually of unregulated electric energ Scotland's hydroelectric potential is about doubl this. *See* ENERGY SOURCES; WIND. [E AY.

Wind rose

A diagram in which statistical information concern ing the direction and speed of the wind at a par ticular location may be conveniently summarized In the standard wind rose a line segment is draw in each of perhaps eight compass directions from common origin The length of a particular segmen



Standard wind rose

is proportional to the frequency with which winds blow from that direction Parts of a given segment are given various thicknesses, indicating frequencies of occurrence of various classes of wind speed from the given direction. *See* WIND [F.S.]

Wind stress

The drag or tangential force per unit area exerted on the surface of the earth by the adjacent layer of moving air. Erosion of ground surfaces and the production of waves on water surfaces are manifestations of wind stress. Surface wind stress determines

the exchange of momentum between the earth and the atmosphere and, together with internal atmospheric viscous stresses, exerts a strong influence on the typical variation of wind through the lowest few thousand feet of the atmosphere. Estimated values of the surface wind stress range up to several dynes/cm², depending upon the nature of the surface and upon the character of the adjacent air flow

Internal horizontal stresses. Significant stresses arise within the lower atmosphere because of the strong shear of the wind between the slowly moving air near the ground and the more rapidly moving air a few thousand feet above

be dimensions ranging up to a few hundreds of meters. The effectiveness of this turbulent viscosity in transferring momentum from level to level may in favorable circumstances be 1,000,000 times as great as the effectiveness of purely molecular viscosity These turbulent viscous stresses exert an indirect effect on the surface wind stress by effecting the transfer of large amounts of momentum from higher levels down to levels adjacent to the surface of the earth.

The torque exerted by the surface wind stress averaged over the entire earth and over a sufficiently long period of time must be equal to zero. Otherwise the net torque acting between the earth and the atmosphere would tend to alter the rate of rotation of the earth about its axis and to alter the mean wind circulation of the atmosphere In the atmosphere the angular momentum supplied, in effect, by the earth in regions of easterly surface flow balances the angular momentum drained by the earth in regions of westerly wind flow, so that the average torque exerted by the surface wind stress is indeed zero, or very nearly so The average magnitude of the surface wind stress is, of course, not equal to zero It determines the average rate at which kinetic energy of the winds is dissipated by surface friction.

Wind pressure. This is the force exerted by the wind per unit area of solid surface exposed normal to the wind direction In contrast to shearing stresses, the wind pressure arises from the difference in pressure between the windward and lee sides of the exposed surface. Wind pressure thus represents a substantial force when the wind speed is high. See GEOSTROPHIC WIND; METEOROLOGY, OCEAN CURRENTS; WIND. [F.S.]

Bibliography: B Haurwitz, *Dynamic Meteorology*, 1941, O G. Sutton, *Micrometeorology*, 1953

Wind tunnel

A duct in which the effects of air flow over objects can be determined The basic law under which wind tunnels operate is that forces on a body held still in moving air are the same as those when the body moves and the air is still, speeds and air conditions being held constant The advantages of holding the model still are that measurement, instrumentation, and flow observation are made much more easily and propulsion is not required. Design and construction of a fixed model is simpler than of a flying model In most instances, for reasons of economy, models (rather than full-scale objects) are tested in wind tunnels, and appropriate scaling laws are employed to extend the data to full scale.

Related research equipment with which aerodynamic studies are accomplished include shock tubes, ballistic guns, drop models, rockets, whirling arms, hot shot tunnels, rocket-powered sleds, and flight tests.

Test conditions. Wind tunnels are used primarily for research and development of vehicles that travel through the air Typical problems suitable for solution in wind tunnels are listed below. The tunnels in which experiments are made should closely match the conditions of full-scale flight (*see* SIMILITUDE). However, vehicular speed and size do not ordinarily require duplication The two parameters one seeks to match are scale and compressibility effects.

Determinations made with wind tunnels

Drag of airplanes and missiles for performance calculations

Level of static and dynamic stability of airplanes and missiles

Pressure distribution around and hence airloads on airplanes or missiles to aid in their structural design

Distribution and rate of heat transfer to airplanes or missiles for cooling and structural design

Torques required to deflect control surfaces

Performance of air-breathing engines and of the air flow into them

Performance of propellers

Lifting characteristics of winged vehicles

Conditions for release and separation of bombs and missiles from aircraft

Effect of winds on buildings, bridges, light poles, signs, and other nonflying structures

Nature of the flow of smoke or other combustion products from smokestacks

Nature of wind flow around geographic features

Scale is duplicated when both model tests and full-scale flight are at the same Reynolds number where Reynolds number is the ratio of inertia forces of the air to viscous forces in the air Disregarding compressibility effects, equal Reynolds number means that the flow patterns about model and full scale are similar *See* REYNOLDS NUMBER

Compressibility effects are similar when the ratio of relative air velocity to the local speed of sound in the air is duplicated *See* MACH NUMBER.

Speed of sound is rarely measured or controlled in the tunnel. it turns out that Mach number is both easy to measure and adjust directly

In general, at speeds below a Mach number (M) of 0.7, test Reynolds numbers of 1,500,000 and up allow adequate duplication of flow over a full-scale model in flight, the speed or Mach number at which the test is run is inconsequential. Above

about $M = 0.7$ air becomes markedly compressible and the primary parameter that must be matched is the Mach number of flight. With compressible flow, Reynolds numbers of 4,000,000 and up are desirable but not mandatory.

Data taken from models in wind tunnels invariably suffer from the extraneous effects on the airstream of the model supporting structure. These effects can be reduced by carefully arranging and minimizing the support structure, and by employing several variations in supports so that an extrapolation to zero support may be attempted.

Measurements. Most data are secured in wind tunnels through one of five methods: (1) measurements of model surface pressures, (2) measurements of model forces and moments, (3) measurements of changes produced in the airstream by the model, (4) measurements of local temperatures, and (5) visual studies. *See also* WIND TUNNEL INSTRUMENTATION.

Surface pressures Surface pressures may be measured by connecting orifices flush with the model surface to pressure measuring devices. The information so obtained will yield local air load and hence the necessary structural strength The information also provides data for shape variation (streamlining) studies and heat transfer calculations, and total surface forces and moments, through integration of pressures.

Pressure gradients for boundary layer type determination are also obtained. The boundary layer arises because the air next to a surface is slowed by viscous action, until at the surface its velocity is zero. Boundary layers vary in thickness from a fraction of an inch on models to a few inches on full-scale airplanes. The boundary layer thickens as one proceeds downstream, and thins with cooling or increased Reynolds number

Forces and moments Force and moment measurements are made through the use of balances, devices which separate the forces along the three axes (lift, drag, and side force) and the moments about the three axes (yawing, rolling, and pitching). The point about which moments are taken is selected in the balance design. Balances may be external to the wind tunnel with the forces and moments brought to them through support struts, or internally in the model.

Reaction of model on the airstream. Because the force of the airstream on the model is equal but opposite to the force of the model on the airstream, measurements of the stream changes produced by the model may be interpreted as the forces and moments on the model. The change of flow inclination as indicated by the pressure changes on the tunnel floor and ceiling may, in some tunnels, be reduced to the lift and pitching moment on the model. The loss of momentum of the airstream as it passes the model indicates a portion of the model's air resistance

Surface temperatures. Measurements of surface temperatures indicate the amount of heat being transferred from the airstream to the model or vice versa, the amount of cooling that must be provided for vehicle strength and comfort, and the type of boundary layer flow

Flow visualization. At velocities near or above the speed of sound, a visible indication of some properties of the flow may be obtained through the use of optical devices (*see* INTERFEROMETRY; SCHLIEREN PHOTOGRAPHY; SHADOWGRAPH OF FLUID FLOW). The Schlieren method has a light cancellation arrangement such that only light deflected by changes in density within the test section is visible. It is used primarily for qualitative flow measurements such as the location of changes in the flow close to the model or of the shock waves The shadowgraph reveals changes in density through the principle that light is refracted due to variations in density of the flow medium. It is simpler

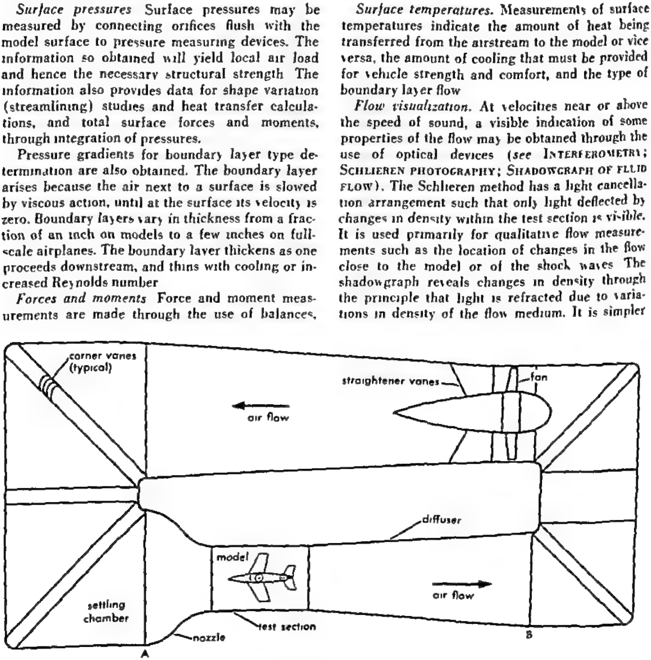


Fig. 1. Outline plan view of low-speed return-flow wind tunnel.

than a Schlieren, but only useful for gross changes. The interferometer depends on phase interference which produces interference fringes To a degree it may be used to make approximate quantitative measurements of pressure variations.

Effects of the model on the flow may be made visible by introducing smoke filaments into the airstream, affixing tufts to the model surface, or by coating the model with a mixture of lampblack and kerosine which then shows surface streamlines as streaks of lampblack.

Low speed wind tunnels. A typical low-speed wind tunnel has a device to drive the air, a design such that smooth flow results in a zone (called the test section) in which the test objects are employed, and instrumentation to measure the effects of air flow over the test object (Fig. 1). A closed-throat return flow tunnel uses the same air over and over again; the tunnel may run continuously. The same test section (Fig 1) could operate if the fan were moved to B and only the circuit between A and B were used, the rest being removed. The purpose of the return section is to ensure smooth flow and to avoid extraneous air currents; its use does not conserve power. The shape of the tunnel, continuously expanding from test section past fan to settling chamber, reduces losses at the corners and return passage by reducing the air velocity through them and improves the test section flow by making possible a contracting entrance to it.

The propeller (more properly called a fan in this application) causes an increase of static and stagnation pressure; there is no increase of velocity through the confined fan as in the case of a free airplane propeller. Static pressure is the pressure measured by a barometer moving with the airstream Stagnation pressure is that measured when the stream is brought to rest; it is analogous to reservoir height in hydraulics

Low speed tunnels are satisfactory up to speeds of about 300 miles per hour in the test section. Above that speed, the power required is such that an air cooler (usually placed upstream of the settling chamber) is required to keep the temperature of the tunnel low An alternate to a cooler is to provide louvers through which the hot air can be dumped and cool outside air brought in Air exchanged in this manner is typically 10–20% of the rate of flow through the circuit

An approximation of the power required for a low speed wind *tunnel* may be made from

$$\text{Horsepower required} = 1.5\ A_{ts}(v/100)^3(d/0.076)$$

where A_{ts} is test section area in ft², v is air velocity in mph, d is air density in lb/ft³ (standard sea level air having a density of 0.076 lb/ft³). That is, a low-speed tunnel requires about 1.5 horsepower per square foot of test section when the air speed is 100 mph and the air density is the standard sea level value

The raw data from model tests in a low-speed closed-throat tunnel are erroneous in several respects Forces are too high because the walls restrict the natural spread of the air about the model. This effect is called solid blocking, and is typically 1% or so. The forces are also too high because of a restriction of the flow behind the model. This effect is called wake blocking and is also typically about 1%.

Measured lift on lifting models, such as winged missiles or aircraft, is too high at a particular angle of inclination to the stream because the solid surfaces of the test section restrict downflow. This effect is resolved by use of a downwash correction which is typically less than 10% of the wing angle of inclination. A smaller correction of similar nature is called streamline curvature correction.

If the velocity in the test section of a wind tunnel increases toward the downstream end, due to the growth of boundary layers along the walls, the force in the downwind direction (called drag) will be too high This effect is called horizontal buoyancy and is typically less than 1%

Because of the small scale of typical models, the maximum force perpendicular to the airstream (called lift) will be smaller and the parallel force (drag) will be larger than full scale The error (called Reynolds number effect) can be substantial, and is not subject to simple analysis Also, turbulence in a tunnel requires thorough study

If the walls of the test section are removed to form a free jet, the above corrections are changed. Solid blocking becomes smaller and of opposite sign, wake blocking and buoyancy become negligible, and downwash becomes larger and of opposite sign With the exception of the scale effect, the corrections to data taken in a wind tunnel are well understood

Transonic tunnels. A conventional low-speed tunnel, even with the addition of a cooler, cannot be driven above the speed of sound without extensive

duct only at a place where the duct area is a minimum. The zone of minimum cross sectional area of a low-speed tunnel is the test section (or more accurately the part of the test section where the model is placed), hence sonic speed will occur at this point. With sonic flow at the model, additional power will only extend the zone of supersonic flow downstream of the model

To achieve transonic flow, the test section must be surrounded by and vented to a plenum tank. Vents in the test section walls may be slots (comprising 10% of the wall surface) parallel to the air flow or perforations of about 20% of the surface area The purpose of the ventilations is as follows. Shock waves (waves in the airstream through which the velocity suddenly decreases, and static pressure and entropy increase) appear in the vicinity of every model when the airstream approaches the speed of sound ($M = 1.0$). The nature of these waves is such that they are reflected unchanged from a solid boundary but with a change in sign

from a free boundary. By providing sufficient free surface in the form of perforations or slots, reflections which would alter the flow over the model are canceled. Ventilation is useful in the range of testing speeds from $M = 0.8$ to $M = 1.4$, and even higher for some long-body tests

The raw data from ventilated wind tunnels are directly applicable with the exception of scale effects. Any wall effects that exist are reduced by the practical limitation that models tested in transonic wind tunnels are necessarily small relative to the size of the test section.

Transonic tunnels usually require about 500 hp per square foot of test section This power may be reduced if the tunnel pressure and hence the air density is lowered.

Supersonic wind tunnels. For test speeds greater than about $M = 1.4$, several changes in tunnel configuration are necessary. The most profound difference between the supersonic and lower-speed tunnels is that a constriction or throat must be provided upstream of the test section in order for the flow to become supersonic This requirement arises through the previously mentioned condition that sonic speed can be obtained only at a minimum section. Thus, with a throat upstream of the test section, the flow accelerates subsonically to the throat, reaches sonic speed at the throat, and accelerates to supersonic flow as the passage expands toward the downstream end. Once sonic speed has been obtained at the throat, the Mach number at any point downstream is uniquely determined by the ratio of the area at the point in question to that of the throat (Fig 2).

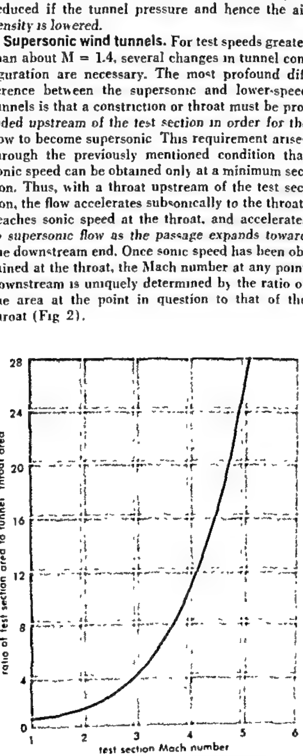


Fig 2. Ratio of test section area to throat area for supersonic air velocities (Mach numbers above 1).

Continuous supersonic wind tunnels (depending on their speed and pressure capabilities) typically require from 2000 to 5000 hp per square foot of test section. This tremendous power has led to the development of tunnels designed to store energy at a low rate in the form of pressure or vacuum (or both) and to discharge at about 50 times the storing rate The reduction in cost of such an intermittent tunnel plus the success of moderately rapid instrumentation has resulted in large numbers of intermittent tunnels being built They are arranged in a majority of the installations to operate on the blow-down principle, exhausting high pressure air through a pressure regulator and duct system to the atmosphere (Fig. 3).

Another system called an indraft tunnel permits air to stream through the duct system into a vacuum tank (Fig. 4) Advantages of the indraft tunnel over the blow-down type include constant stagnation pressure and stagnation temperature, greater safety, ease of operation, and less noise. Advantages of the blow-down type over the indraft tunnel include higher Reynolds number and ability to vary Reynolds number over a wide range.

There are two major problems in the design of supersonic wind tunnels. The first is to supply sufficient pressure ratio between the settling chamber and the air exit to start and sustain the flow either by compressors or by storing air at adequate pressure (Fig. 5) The second problem is to supply air dry enough to avoid excessive moisture condensation during the expansion process in the supersonic nozzle (and subsequent ragged flow in the test section) through various air drying procedures, usually resulting in a dew point at atmospheric pressure of about −40°F

Data taken in a supersonic wind tunnel may be used directly with only allowances for scale effects and (rarely) a small correction for horizontal buoyancy

Hypersonic wind tunnels. Hypersonic wind tunnels are those whose speed range is from about $M = 5$ to $M = 15$ (speeds above $M = 15$ are extraordinarily difficult to develop)

Although there are a few continuous hypersonic wind tunnels, the majority are either of the blow-down or combined blow-down and indraft type. The combined tunnel is especially suitable for developing the exceedingly high pressure ratios needed to start and run (Fig 6)

In general hypersonic tunnels have the same design features as supersonic wind tunnels with the following additional features To provide the exceedingly high pressure ratio either a large number of series-staged compressors or a blow-down and indraft arrangement must be provided. Because the drop of stream temperature as the air expands in the supersonic portion of the nozzle may carry the air below its liquefaction temperature, additional heat may need to be provided This is usually accomplished by passing the air through a gas-fired or electrical-resistance heater. The nozzle may require cooling. The tremendous expansion of the air as it proceeds through the super

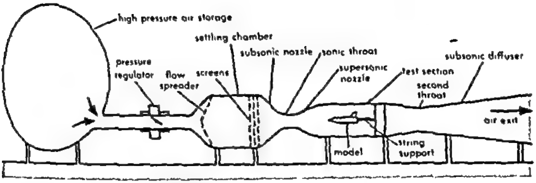
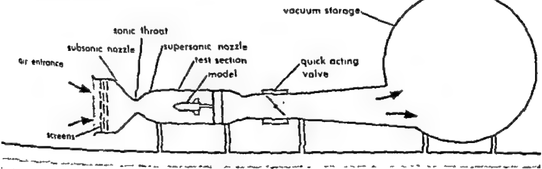


Fig 3. General configuration of an intermittent supersonic blow-down wind tunnel



ig 4 Elements of an intermittent indraft supersonic ind tunnel

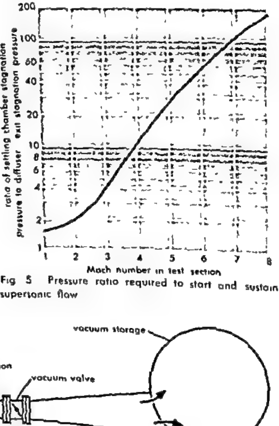
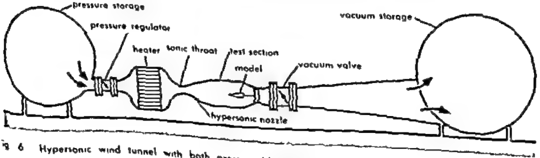


Fig 5 Pressure ratio required to start and sustain supersonic flow

nic nozzle results in a low density in the test ction, which in turn means a low Reynolds number and a thick boundary layer This in turn complicates the correction for the boundary layer uring the design of the nozzle It also results in elatively large changes in Mach number with hanges in stagnation pressure because large hanges in the boundary layer ensue

Data from hypersonic wind tunnels are directly pplicable (with perhaps a small correction for orizontal buoyancy) to free stream conditions here may be serious Reynolds numbers effects if e test Reynolds numbers are too far from full cale *See* AIRCRAFT TESTING, MODEL THEORY, OCKET SLED TESTING. [A PO.]

Bibliography. R C. Pankhurst and D W Holden, *ind Tunnel Technique*, 1952, A Pope, *Wind Tunnel Testing*, 2d ed., 1954.



ig 6 Hypersonic wind tunnel with both pressure blow-down and vacuum indraft

Wind tunnel instrumentation

Measuring devices used in wind tunnel tests. Wind tunnel instrumentation requires, in addition to conventional laboratory instruments for fluid flow, thermometry, and mechanical measurements, sensing devices capable of precision measurement in the small-scale environment of the test setup. As with other research instrumentation, equipment of greater sensitivity and faster response than is used for normal operation is also desired. The principal dimensions measured are humidity, temperature, turbulence, direction of flow, and mass flow.

Humidity measurement. Moisture content of the air in a wind tunnel is important; at subsonic velocity it affects the apparent density of the air, and at supersonic velocity it may cause condensation shocks which will result in a variation of Mach number in the test section. The conventional methods such as the wet- and dry-bulb method are not satisfactory because the minimum dew point for the small supersonic wind tunnel is 0°F or less, and for the larger supersonic tunnel it is −40°F or less for satisfactory operation.

The technique used at the NASA Lewis Research Center since 1942 is as follows. The sample to be tested is pumped from the tunnel stream into a chamber at a pressure above atmospheric and then rapidly expanded to atmospheric pressure. The resulting expansion and cooling of the sample, if sufficient, will condense water vapor out as a fog. The point where charging pressure is high enough for expansion to atmospheric pressure to be just sufficiently cooling to condense the water vapor to a fog is the dew point of the sample. The process is adiabatic, so that the initial temperature of the sample and the expansion ratio are all that are needed to give the dew point.

Temperature probes. In low subsonic tunnels the total temperature and the static temperature are close, but in the higher subsonic and supersonic tunnels the static temperature is much lower than the total temperature. The temperature usually recorded is the total temperature or stagnation temperature of the flow; the static temperature is computed from this measurement.

For the average temperature measurements required in wind tunnel tests, the bare-wire thermocouple is sufficient (*see* TEMPERATURE MEASUREMENT). However, where the wall temperature is greatly different from the gas temperature, a shielded thermocouple must be used. In the case of high-temperature tunnels, such as rocket or combustion, special sonic-flow temperature probes must be used.

Turbulence measurements. Turbulence is characterized by a time variation in velocity and pressure or temperature (*see* TURBULENT FLOW). To study turbulence an instrument must be able to indicate the time variations of the motion.

The hot-wire anemometer is a transducer which gives an electrical signal proportional to the flow fluctuations. The sensing element of the hot-wire anemometer is a circular cylinder 0.00005–0.0005 in. in diameter made of platinum or tungs[ten] heated electrically to a temperature greater th[an] the surrounding fluid. The amount of heat c[on]vected away from the cylinder surface is prop[or]tional to the flow velocity, density, and tempe[ra]ture.

The hot-wire anemometer has a finite mass a[nd] cannot accurately reproduce fluctuations grea[ter] than about 100 cycles per second. For frequenc[ies] higher than this it is necessary to operate the wi[re] with an auxiliary electronic circuit to compens[ate] for the thermal lag of the wire. This circuit is [re]ferred to as a constant-current anemometer. T[he] hot wire is operated as one arm of a Wheatsto[ne] bridge, because a change in wire temperature [is] directly related to a change in wire resistance. T[he] output of the wire is fed into a compensated amp[li]fier, which is adjusted to the correct compensati[on] by electronic calibration of the wire.

The anemometer system described above h[as] been used extensively in turbulent research, but [is] now being replaced by the more sophisticat[ed] direct-current feedback, constant-temperature an[e]mometer system. The amplifier is a servomech[a]nism, which senses an unbalance in the Wheatsto[ne] bridge due to a change in the hot-wire resistan[ce] and feeds back a current to restore the wire to th[e] resistance necessary to balance the bridge. In th[is] way the wire is held at a constant instantaneou[s] resistance, and thus a constant temperature, regar[d]less of the amount of heat being transferred fro[m] the wire. With the wire held at a constant instan[taneous] temperature no thermal lag occurs, an[d] the feedback current is a measure of the fluctuatin[g] turbulence.

The very small diameter hot wires are subject t[o] breakage, thus a hot-film anemometer has been sug[gested] as an alternative transducer. A very thi[n] film of metal, usually platinum, is coated on a[n] insulator and the film replaces the wire.

The use of dye traces in liquids is perhaps th[e] oldest method of studying turbulent motion and i[s] of great value in understanding turbulent motion.

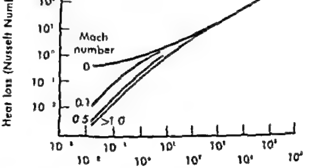


Fig. 1. Heat loss from a circular cylinder normal to the flow

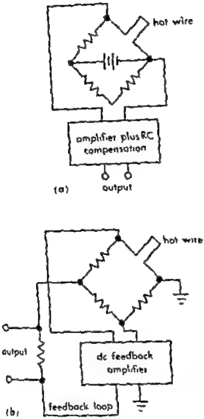


Fig 2 Hot-wire anemometer (a) Constant-current. (b) Constant temperature

Schlieren photos allow the turbulent flow to be visualized in supersonic air streams (*see* SCHLIEREN PHOTOGRAPHY).

Indirect methods in wide use for indicating the degree of turbulence in wind tunnels are the measuring of the drag of spheres in subsonic flow and the recovery temperature along cone surfaces in supersonic flow.

Direction probes. The flow direction and the degree to which the flow is parallel to the tunnel walls is vital in the analysis of pressure and forces on a test model in a wind tunnel Several devices are used the type depending on the flow velocity (Fig 3)

Yaw sphere In subsonic flow, the pressure distribution on the surface of a sphere can be used to determine the flow direction of the air passing over the sphere In a typical probe, yaw angle as well as pitch angle can be determined simultaneously The ratio of the difference over the average of opposite static pressures should be calibrated against the flow angles before the probe is used in a tunnel.

Wedge In supersonic flow the sphere is not usable, instead the wedge serves in its classical use as a means to determine the velocity of a supersonic stream and as an angle-of-flow meter. The wedge is a two-dimensional device, so that only either pitch or yaw direction can be measured at a time If the wedge is used only for directional measurements the total pressure tubes are not needed if the Mach number is known The angle of flow can be obtained from tabulated data for wedges in supersonic flow. The wedge with static pressure tubes only can be used to indicate subsonic flow angles but must be calibrated before it is used.

Cone. Similar to the wedge, the cone is usable in both supersonic and high subsonic flow fields to determine flow angles. The cone can be used to measure yaw and pitch angles simultaneously. However, like the sphere, it must be calibrated.

[illegible] When space is not restricting the mechanica[illegible] yaw or pi[illegible] vane is mounted at its center of gravity on a calibrated position transducer and is read remotely.

Mass flow. Subsonic wind tunnel mass flow may be metered in any section where the flow conditions can be determined. When total pressure, temperature, static pressure, and area of the section are known, the mass flow is readily calculated. The problem in any mass-flow measurement is to obtain the true average values of the variables. Therefore a detailed survey is usually needed to calibrate a tunnel

Supersonic wind tunnel mass-flow measurements are somewhat easier. The total pressure and temperature of the plenum chamber, with the area of

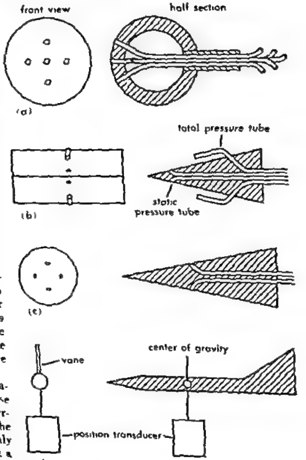


Fig. 3 Direction measuring instruments. (a) Yaw sphere (b) Wedge (c) Cone. (d) Vane.

Wind tunnel instrumentation

Measuring devices used in wind tunnel tests. Wind tunnel instrumentation requires, in addition to conventional laboratory instruments for fluid flow, thermometry, and mechanical measurements, sensing devices capable of precision measurement in the small-scale environment of the test setup. As with other research instrumentation, equipment of greater sensitivity and faster response than is used for normal operation is also desired. The principal dimensions measured are humidity, temperature, turbulence, direction of flow, and mass flow.

Humidity measurement. Moisture content of the air in a wind tunnel is important; at subsonic velocity it affects the apparent density of the air, and at supersonic velocity it may cause condensation shocks which will result in a variation of Mach number in the test section. The conventional methods such as the wet- and dry-bulb method are not satisfactory because the minimum dew point for the small supersonic wind tunnel is 0°F or less, and for the larger supersonic tunnel it is −40°F or less for satisfactory operation.

The technique used at the NASA Lewis Research Center since 1942 is as follows. The sample to be tested is pumped from the tunnel stream into a chamber at a pressure above atmospheric and then rapidly expanded to atmospheric pressure. The resulting expansion and cooling of the sample, if sufficient, will condense water vapor out as a fog. The point where charging pressure is high enough for expansion to atmospheric pressure to be just sufficiently cooling to condense the water vapor to a fog is the dew point of the sample. The process is adiabatic, so that the initial temperature of the sample and the expansion ratio are all that are needed to give the dew point.

Temperature probes. In low subsonic tunnels the total temperature and the static temperature are close, but in the higher subsonic and supersonic tunnels the static temperature is much lower than the total temperature. The temperature usually recorded is the total temperature or stagnation temperature of the flow; the static temperature is computed from this measurement.

For the average temperature measurements required in wind tunnel tests, the bare-wire thermocouple is sufficient (*see* TEMPERATURE MEASUREMENT). However, where the wall temperature is greatly different from the gas temperature, a shielded thermocouple must be used. In the case of high-temperature tunnels, such as rocket or combustion, special sonic-flow temperature probes must be used.

Turbulence measurements. Turbulence is characterized by a time variation in velocity and pressure or temperature (*see* TURBULENT FLOW). To study turbulence an instrument must be able to indicate the time variations of the motion.

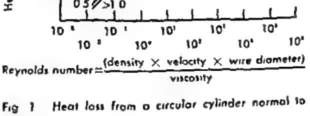
The *hot-wire anemometer* is a transducer which gives an electrical signal proportional to the flow fluctuations. The sensing element of the hot-wire anemometer is a circular cylinder 0.00005–0.0005 in. in diameter made of platinum or tungsten heated electrically to a temperature greater than the surrounding fluid. The amount of heat convected away from the cylinder surface is proportional to the flow velocity, density, and temperature.

The hot-wire anemometer has a finite mass and cannot accurately reproduce fluctuations greater than about 100 cycles per second. For frequencies higher than this it is necessary to operate the wire with an auxiliary electronic circuit to compensate for the thermal lag of the wire. This circuit is referred to as a constant-current anemometer. The hot wire is operated as one arm of a Wheatstone bridge, because a change in wire temperature is directly related to a change in wire resistance. The output of the wire is fed into a compensated amplifier, which is adjusted to the correct compensation by electronic calibration of the wire.

The anemometer system described above has been used extensively in turbulent research, but is now being replaced by the more sophisticated direct-current feedback, constant-temperature anemometer system. The amplifier is a servomechanism, which senses an unbalance in the Wheatstone bridge due to a change in the hot-wire resistance and feeds back a current to restore the wire to the resistance necessary to balance the bridge. In this way the wire is held at a constant instantaneous resistance, and thus a constant temperature, regardless of the amount of heat being transferred from the wire. With the wire held at a constant instantaneous temperature no thermal lag occurs, and the feedback current is a measure of the fluctuating turbulence.

The very small diameter hot wires are subject to breakage, thus a hot-film anemometer has been suggested as an alternative transducer. A very thin film of metal, usually platinum, is coated on an insulator and the film replaces the wire.

The use of dye traces in liquids is perhaps the oldest method of studying turbulent motion and is of great value in understanding turbulent motion.



Reynolds number = (density × velocity × wire diameter) / viscosity

Fig. 1. Heat loss from a circular cylinder normal to the flow.

ructure In the case of dc motors and generators ...th compensating windings, both forms of construction may be found in one machine

Concentrated field windings for salient-pole ac synchronous machines and for shunt fields of dc and series ac machines are form-wound of many turns of insulated copper wire. The completed coil is taped and impregnated to hold its shape and to fit around the pole core Series field coils are normally concentrated They are wound of larger wire, or of insulated copper strip, in order to carry the armature current without overheating

Field coils for cylindrical construction are made of insulated rectangular copper or aluminum bars Individual coils are form-wound with several turns per coil. The dimensions of the coils are selected to provide a suitable space distribution of field flux The coil is held in the slots by nonmagnetic wedges The coil ends are held in place against mechanical forces by nonmagnetic bands

Rotor field coils are connected to the external circuit through slip rings, on which carbon brushes rest to conduct the exciting current *See* SLIP RINGS

Armature windings. These windings carry ac current. They take many forms, depending upon the type and capacity of the machine and specific design requirements Armature windings have an active portion, which lies in slots in the magnetic circuit, and end turns and end connections, which are external to the airgap The treatment of the end connection determines the appearance and operating characteristics of the winding Usually the end turns are formed to make diamond shaped coils for ease in inserting and bracing the windings A coil may have a single turn, or it may have many turns Each turn may have a single strand or it may have several strands of copper electrically in parallel A coil has two coil sides, which lie in slots approximately a pole pitch apart for dc machines, whereas most ac machines have short pitch windings in which the coil pitch is less than 180 electrical degrees. The pitch of a coil is found by taking the quotient of the slots separating the coil sides and the slots in 180 electrical degrees Windings may be single layer, one coil side per slot, or double layer, two coil sides per slot, in which case one side of each coil will rest in the bottom half of its slot and the other coil side at the top of its slot

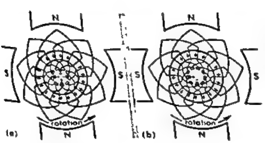


Fig 3 (a) Simple 4-pole lap winding (b) Simple 4-pole wave winding

DC armature windings. These may be lap, or parallel, windings, or they may be wave, or series, windings. Figure 3 shows a 4-pole lap dc armature winding and a 4-pole wave dc armature winding The two positive brushes of the lap winding would be connected together and the two negative brushes would be connected together The lap and wave windings have the following general properties:

1 Coil ends of lap windings are fastened to adjacent commutator bars Coil ends of wave windings are fastened to commutator bars approximately two pole pitches apart

2 The lap winding has as many parallel paths between the line terminals as there are poles, and the conductors of a path lie under adjacent poles. The wave winding has two parallel paths between the line terminals regardless of the number of poles, and conductors of either path lie under all poles.

3 The lap winding has as many sets of brushes as there are numbers of poles, whereas wave windings need only two brush sets Usually, however, the machine carries as many brush sets as there are poles to provide greater current carrying capacity for the commutator Brush positions are such that the coils are short-circuited only when they are in the position of minimum induced voltage.

Lap windings are adapted to high-current machines because they may have more than two parallel paths, whereas the wave windings are adapted to small-capacity machines and high-voltage machines because of the series connection of the coils.

AC armature windings These may be single or polyphase, full or fractional pitch, Y-connected or delta-connected Practically all except small-capacity ac machines have three-phase windings, with coils distributed around the entire armature periphery for better utilization of space and material in the machine. Winding factors are used to evaluate the performance of a winding The generated rms voltage per phase is $E = 4.44\, k_d k_p f N_{ph} \Phi$ where k_d is the distribution factor, k_p the pitch factor, f the frequency in cps, N_{ph} the total turns in series per phase, Φ the flux per pole in webers per square meter

The distribution factor k_d of a winding gives the ratio of the voltage output of a distributed winding to the voltage output of a concentrated winding

$$k_d = \frac{\sin (n\gamma/2)}{n \sin (\gamma/2)}$$

where n is number of slots over which one phase is distributed and γ is electrical angle between slots. Fractional-pitch windings are employed to reduce the magnitude of harmonics, to improve wave shape to save copper and to permit end-connections between coils to be made more easily

The pitch factor k_p of a winding gives the ratio of the voltage output of a fractional-pitch coil to the voltage output of a full-pitch coil.

$$k_p = \cos \frac{\pi - \rho}{2}$$

the minimum section of the supersonic nozzle, give a unique solution. The method is referred to as a choked-orifice method. The minimum section of the nozzle section is sonic, and thus the conditions of the sonic section of the nozzle are known from continuity flow equations.

In tunnels of 12-in. test sections or less the ASME standard orifices may be used for metering air flow. The rounded flow nozzle or the venturi nozzle are the most applicable. *See* AIRCRAFT TESTING; *see also* NOZZLE; PITOT TUBE; VENTURI TUBE. [R.B.C.]

Bibliography: Ames Research Staff, *Equations, Tables, and Charts for Compressible Flow*, NACA Rept. 1135, 1953; *Fluid Meters, Their Theory and Applications*, ASME Research Publ., 4th ed., 1937; L. S. G. Kovasznay, *Technique for the Optical Measurement of Turbulence in High Speed Flow*, ASME, 1949; S. C. Ling and B G. Hubbard, The hot-film anemometer, a new device for fluid mechanics research, *J. Aeronaut. Sci.*, 23(9):890–891, 1956; G. B. Schubauer and P S Klebanoff, *Theory and Application of Hot-Wire Instruments in the Investigation of Turbulent Boundary Layers*, NACA WR W-86, 1946, F. S. Simmons and G. E Glawl, *Theory and Design of a Pneumatic Temperature Probe and Experimental Results obtained in a High-Temperature Gas Stream*, NACA TN 3893, 1957.

Wind-shift line

A line along which the wind veers abruptly. In the Northern Hemisphere, a shift from easterly or southeasterly winds to southerly or southwesterly is normally observed on passage of a warm front, while a shift from southerly or southwesterly to westerly or northerly winds is usually found on passage of a cold front, or a squall line (*see* SQUALL) Passage of wind-shift lines is commonly accompanied by marked changes in general weather conditions (*see* STORM). Lowest barometric pressure is generally found on passage of a cold-front type of wind-shift line. [C W.N.]

Windings (electric machinery)

Windings can be classified in two groups: armature windings and field windings. The armature winding is the main current-carrying winding in which the emf or counter-emf of rotation is induced The current in the armature winding is known as the armature current. The field winding produces the magnetic field in the machine. The current in the field winding is known as the field or exciting current

The location of the winding depends upon the type of machine. The armature windings of dc motors and generators are located on the rotor, since they must operate in conjunction with the commutator, and the field windings are mounted on stator field poles (*see* DIRECT-CURRENT GENERATOR; DIRECT-CURRENT MOTOR). Alternating-current synchronous motors and generators are normally constructed with the armature winding on the stator and the field winding on the rotor. There is no clear distinction between the armature and field windings of ac induction motors or generators. One winding may carry the main current of the machine and also establish the magnetic field. It is customary to use the terms stator winding and rotor winding to identify induction motor windings. The word armature, when used with induction motors, applies to the winding connected to the power source (usually the stator). *See* ALTERNATING-CURRENT GENERATOR, ALTERNATING-CURRENT MOTOR

Field windings. Field windings produce a magnetic pole fixed in space with respect to the magnetic structure on which they are mounted. If salient-pole construction is employed, the winding turns are concentrated around the pole core. If cylindrical construction is employed, the field winding is constructed of elongated concentric loops embedded in slots cut in the surface of the field

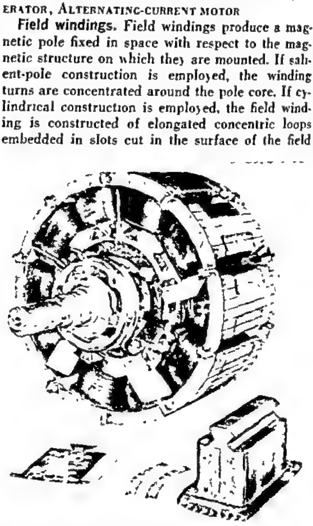


Fig 1 Salient-pole field winding on rotor of synchronous motor (*Allis-Chalmers*)

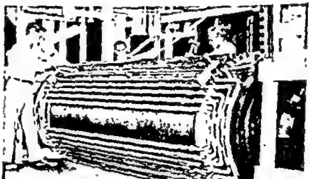


Fig 2. Winding field on cylindrical rotor. (*National Electric Coil Co.*)

tructure In the case of dc motors and generators th compensating windings, both forms of construction may be found in one machine.

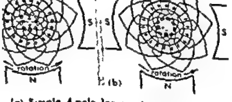
Concentrated field windings for salient-pole ac synchronous machines and for shunt fields of dc and series ac machines are form-wound of many turns of insulated copper wire The completed coil taped and impregnated to hold its shape and to around the pole core. Series field coils are nor- ... wire, ... the

... coils for cylindrical construction are made of insulated rectangular copper or aluminum bars Individual coils are form-wound with several turns per coil The dimensions of the ...

... coil ends are held in place against mechanical forces by nonmagnetic bands.

Rotor field coils are connected to the external circuit through slip rings on which carbon brushes rest to conduct the exciting current *See* SLIP RINGS

Armature windings. These windings carry ac current They take many forms, depending upon the type and capacity of the machine and specific design requirements Armature windings have an active portion, which lies in slots in the magnetic circuit, and end turns and end connections, which are external to the airgap The treatment of the end connection determines the appearance and operating characteristics of the winding. Usually the end turns are formed to make diamond-shaped coils for use in inserting and bracing the windings A coil may have a single turn, or it may have many turns. Each turn may have a single strand, or it may have several strands of copper electrically in parallel. A coil has two coil sides, which lie in slots approximately a pole pitch apart for dc machines, whereas most ac machines have short pitch windings in which the coil pitch is less than 180 electrical degrees. The pitch of a coil is found by taking the ... of the slots separating the coil sides and ... s in 180 electrical degrees Windings may ... le layer, one coil side per slot, or double ... two coil sides per slot, in which case one ... each coil will rest in the bottom half of its ... the other coil side at the top of its slot



(a) Simple 4-pole lap winding (b) Simple 4-... ave winding

DC armature windings. These may be lap, or parallel, windings, or they may be wave, or series, windings. Figure 3 shows a 4-pole lap dc armature winding and a 4-pole wave dc armature winding. The two positive brushes of the lap winding would be connected together and the two negative brushes would be connected together. The lap and wave windings have the following general properties:

1. Coil ends of lap windings are fastened to adjacent commutator bars. Coil ends of wave windings are fastened to commutator bars approximately two pole pitches apart.

2 The lap winding has as many parallel paths between the line terminals as there are poles, and the conductors of a path lie under adjacent poles. The wave winding has two parallel paths between the line terminals regardless of the number of poles, and conductors of either path lie under all poles.

3 The lap winding has as many sets of brushes as there are numbers of poles, whereas wave windings need only two brush sets Usually, however, the machine carries as many brush sets as there are poles to provide greater current carrying capacity for the commutator. Brush positions are such that the coils are short-circuited only when they are in the position of minimum induced voltage.

Lap windings are adapted to high-current ma- ...

chines because of the series connection of the coils

AC armature windings These may be single or polyphase, full or fractional pitch, Y-connected or delta-connected. Practically all except small-capacity ac machines have three-phase windings, with coils distributed around the entire armature periphery for better utilization of space and material in the machine Winding factors are used to evaluate the performance of a winding The generated rms voltage per phase is $E = 4.44\, k_d k_p f N_{ph} \Phi$ where k_d is the distribution factor, k_p the pitch factor, f the frequency in cps, N_{ph} the total turns in series per phase Φ the flux per pole in webers per square meter

The distribution factor k_d of ... ratio of the ... to the voltag...

$$k_d = \frac{\sin (n\gamma/2)}{n \sin (\gamma/2)}$$

where n is number of slots over which one phase is distributed and γ is electrical angle between slots Fractional-pitch windings are employed to reduce the magnitude of harmonics, to improve wave shape, to save copper and to permit end-connec- ...

of ... th... age output of a full-pitch coil

$$k_p = \cos \frac{\pi - \rho}{2}$$

where ρ is electrical angle between coil sides.

Since the angle γ and the angle ρ in the equations for k_d and k_p change with frequency, they may be selected to eliminate or minimize certain harmonic frequencies.

Figure 4 shows the active conductors of a distributed, three-phase, two-pole, fractional-pitch, armature winding wherein a, b, and c are conductors of three phases displaced 120 electrical degrees apart. Coil $a_1, -a_1$ spans ⅚ of a pole pitch, or 150 electrical degrees.

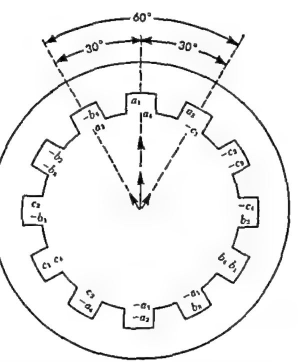


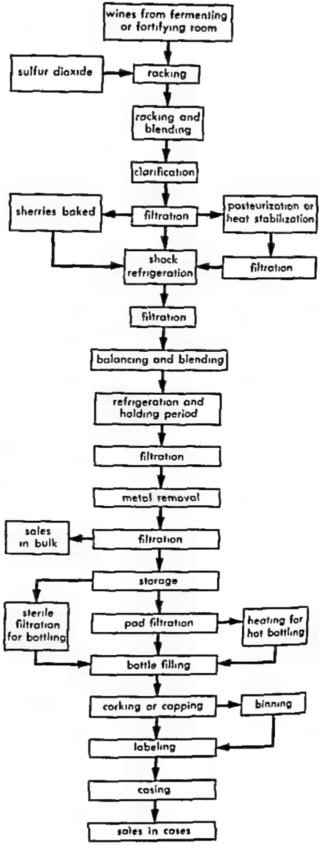
Fig. 4 Distributed 3-phase 2-pole fractional-pitch armature winding with voltage vector diagram (A E Fitzgerald and C. Kingsley, *Electric Machinery*, Mc-Graw-Hill, 1952)

Stator and wound-rotor windings of polyphase induction motors conform to the above descriptions Single-phase induction motor stators may be wound in the same manner, except for the number of phases; or they may be formed of several concentric loops of wire. The windings are usually form-wound from one continuous length of wire, but are not taped. This facilitates the installation of the winding in small, partly closed slots. *See* INDUC-TION MOTOR for additional

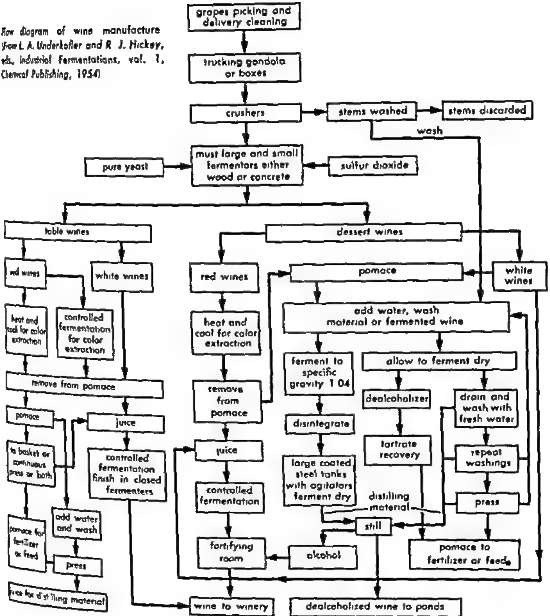
Wine

Alcoholic beverages made by fermentation of the juice of fruits or berries. Wine, as defined by government agencies, is essentially the fermented alcoholic beverage made from grape juice. Wines from other materials are always required to show their source on the label, for example apple wine, berry wine, cherry wine, and the like. California produces over 80% of the grape wine made in the United States, and well over a hundred varieties o the cultivated grape, *Vitis vinifera*, are used.

Classification of wines. This depends on th color, relative sweetness, alcoholic content, pres



Flow diagram of cellar operations in a winery. (L. A Underkofler and R J. Hickey, eds, *Industrial Fermentations*, vol 1, Chemical Publishing, 1954)



Flow diagram of wine manufacture (From L. A. Underkofler and R. J. Hickey, eds., Industrial Fermentations, vol. 1, Chemical Publishing, 1954)

ence of carbon dioxide, the variety of grape, and the region where the grapes are grown. Wines may be red or white. In the production of red wines, the red anthocyanin pigments are extracted from the skins of blue grapes by allowing the fermentation to take place in contact with the skins. When white wines are made from dark grapes, the skins are separated before fermentation. The terms dry and sweet refer to the relative sugar content of a wine. Table wines contain less than 14% alcohol by volume and dessert wines contain over 14%, usually 20% alcohol. The higher alcohol content of dessert wines is obtained by the addition of brandy, which is called fortification. Sparkling wines such as champagne and sparkling Burgundy contain carbon dioxide. Examples of red, dry, table wines are Burgundy, claret, Chianti, cabernet, barbera, and zinfandel. The first three wines are made from various grape varieties, and each has a characteristic color, body, and flavor. The last three are made of the corresponding grape varieties. Some important white table wines are Rhine wines (such as Moselle, Riesling, Traminer), and Chablis, white Chianti, dry sauterne, and sweet sauterne.

Examples of dessert wines are angelica, an amber- or yellow-colored sweet wine without muscat flavor, of California origin, muscatel, a sweet wine made in part from muscat grapes, with a pronounced muscat flavor, port, a deep red color; tawny port, a more amber-tinted sweet wine; white port, a light-colored sweet wine without muscat flavor; sherry, characteristic flavor and aroma and varying sugar content, but less than 7%; and Tokay, a sweet wine with a pink tint.

Wine production. Grapes are harvested when they have reached the desired sugar and acid con-

where ρ is electrical angle between coil sides.

Since the angle γ and the angle ρ in the equations for k_d and k_p change with frequency, they may be selected to eliminate or minimize certain harmonic frequencies

Figure 4 shows the active conductors of a distributed, three-phase, two-pole, fractional-pitch, armature winding wherein a, b, and c are conductors of three phases displaced 120 electrical degrees apart. Coil $a_1, -a_1$ spans $5/6$ of a pole pitch, or 150 electrical degrees.

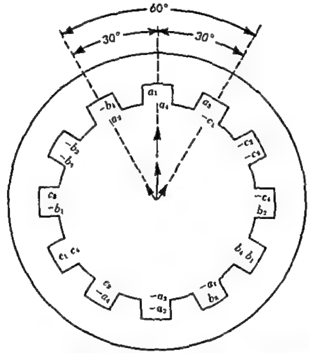


Fig 4 Distributed 3-phase 2-pole fractional-pitch *armature winding with voltage vector diagram (A E Fitzgerald and C Kingsley, Electric Machinery, Mc-Graw-Hill, 1952)*

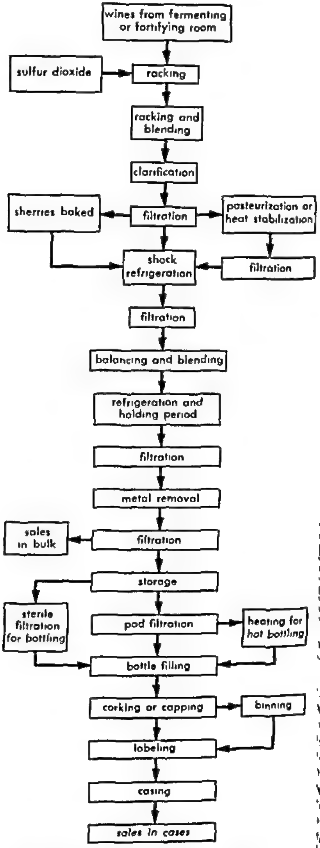
Stator and wound-rotor windings of polyphase induction motors conform to the above descriptions Single-phase induction motor stators may be wound in the same manner, except for the number of phases; or they may be formed of several concentric loops of wire. The windings are usually form-wound from one continuous length of wire, but are not taped. This facilitates the installation of the winding in small, partly closed slots. *See* INDUCTION MOTOR for additional information on these and squirrel-cage windings. [A R.E]

Bibliography: M. Liwschitz-Garik, *Winding Alternating Current Machines*, 1950.

Wine

Alcoholic beverages made by fermentation of the juice of fruits or berries. Wine, as defined by government agencies, is essentially the fermented alcoholic beverage made from grape juice Wines from other materials are always required to show their source on the label, for example apple wine, berry wine, cherry wine, and the like California produces over 80% of the grape wine made in the United States, and well over a hundred varieties of the cultivated grape, *Vitis vinifera*, are used.

Classification of wines. This depends on the color, relative sweetness, alcoholic content, pres-



Flow diagram of cellar operations in a winery. *(L. A Underkofler and R J. Hickey, eds, Industrial Fermentations, vol 1, Chemical Publishing, 1954)*

cially on high-speed airplanes, all structural members may be enclosed within the wing contour. The structure is then an internally braced cantilever. For torsional stiffness, the metal covering serves also as trusses above and below the internal braces, the arrangement being called a semimonocoque structure. In extremely thin wings where space prohibits internal braces, a full monocoque structure is used, the shell carrying the full load.

The wing need not be completely rigid. It may sag when at rest from its own weight or the weight of fuel tanks inside it and engines mounted in nacelles on it, and it may bow upward within its elastic limit from aerodynamic lift while in flight (Fig. 1). However, the aerodynamic properties of the wing should not change significantly in the presence of such deformations. *See* AEROELASTICITY.

When a swept-back wing bends, its angle of attack in the streamwise direction is reduced (Fig. 3). For example, if flight conditions produce an upward force, a wing that is uniformly flexible from leading edge to trailing edge will deflect approximately the same distance along a line normal to its quarter chord. The deflection at the leading edge of a streamwise segment will then be less than at the trailing edge. The angle of attack along a streamwise segment will decrease with consequent decrease in lift. *See* AIRPLANE.

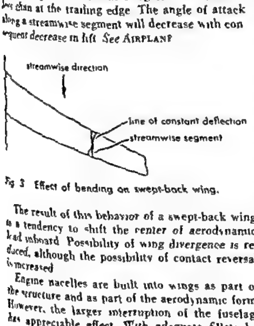


Fig. 3. Effect of bending on swept-back wing.

The result of this behavior of a swept-back wing is a tendency to shift the *center of aerodynamic load* inboard. Possibility of wing *divergence* is reduced, although the possibility of contact reversal is increased.

Engine nacelles are built into wings as part of the structure and as part of the aerodynamic form. However, the larger interruption of the fuselage has appreciable effect. With adequate fillets between wing halves and fuselage, the lift contributed by the fuselage can be almost equivalent to a central wing portion of like width. The drag of wing and fuselage together is considerably more than the sum of drags of the parts in isolation from each other. Wing tips also constitute a discontinuity that is not amenable to simple theoretical analysis. For these reasons exact performance is extrapolated from flight tests of similar vehicles or wind tunnel tests of scaled models.

Movable auxiliary fittings are added to wings so that the pilot can vary their characteristics in flight (Fig. 4). Lift increases at first with increasing angle of attack; however, at wide angles, the air separates from the top of the wing surface with abrupt decrease in lift, a condition called stall. A leading edge slot that can be opened when greater-than-normal lift is needed admits air from the high-pressure region of the lower portion of the leading edge into the low-pressure region above the wing. *See* BOUNDARY-LAYER FLOW.

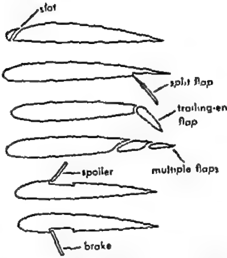


Fig. 4. Auxiliary wing fixtures for controlling characteristics.

Another means for increasing lift is to deflect a trailing-edge flap into the air stream. The flap usually extends across the central portion of the wing leaving the outer portions of the trailing edge for ailerons (see FLIGHT CONTROLS). Both slots and flaps increase drag and so are suited only for use under such extreme conditions as take off.

To limit the speed of an airplane in a dive or to decelerate it during landing, a spoiler can be deflected above the wing or a brake below it. The effects of these devices can be analyzed only approximately. Their exact characteristics are usually determined by experience and test. *See* AIRCRAFT TESTING, AIRFOIL, WING STRUCTURE.

[P H R.]

Bibliography: R. L. Bisplinghoff, H. Ashley, and R. L. Halfman, *Aeroelasticity*, 1955; D. J. Peery, *Aircraft Structures*, 1950; R. von Mises, *Theory of Flight*, 1945.

Wing loading

The ratio of the weight W of an airplane to the area S of its wings. Wing loading W/S is expressed in lb/ft² or kg/m². Slow-speed airplanes have low *wing loading*. High-speed airplanes are heavily loaded, that is, their wing area is relatively small compared to their weight. Wing loading is a design criterion for both the airplane and the airfield. Propeller driven transports designed for minimum operating costs have the highest permissible wing loading. Landing-field d... of wing load... landing stri...

tent and are taken to the winery in boxes or in bulk. Several pickings are required to produce fine wines. The grapes go through a crusher-stemmer, which crushes the berries, but not the seeds, and removes the stems. The must, still containing seeds and skins, is pumped to fermentation tanks where it receives about 100–150 ppm of sulfur dioxide (SO_2) gas. This controls, to a large extent, the growth of bacteria and wild yeasts, whereas the various strains of wine yeast are adapted to this amount of SO_2. A pure yeast starter, consisting of *various strains of Saccharomyces cerevisiae*, is then added. In the production of white wines, the juice is separated from the skins and seeds at an early stage of the fermentation, often after 4–8 hours. The residue may be pressed and extracted for the manufacture of lower quality wines or it may be fermented to completion and used for distillation. Many variations are in use in different wineries, depending on the type of wine to be made. The fermentation tanks, especially the large ones, are cooled to keep the temperature of the wine at 23.9–29.4°C.

After the initial fermentation is over, the wine is transferred to the storage cellar for completion of fermentation, clarification, aging, stabilization, and bottling. These operations are called cellar practices. The sediment of yeast and other insoluble matter is called lees. The new wine is drawn off, or racked, from the lees to avoid picking up undesirable flavors from the lees. Aging is then continued. Wines are often chilled in order to precipitate excess cream of tartar, the potassium acid tartrate from the grape, which otherwise might precipitate upon chilling of bottled wine. Champagne is made by allowing a secondary fermentation to occur with a special flocculating type of yeast, either in bottles, by the original French procedure, or in bulk, using pressure tanks followed by bottling. The CO_2 of sparkling wines is produced by yeast fermentation. Most sherry wines of California obtain their characteristic flavor and aroma by a heating process, called baking, over an extended period of time. Spanish sherries are made by the "flor" process, involving the participation of a film-forming yeast growing for many months at the surface of the wine in partially filled oak barrels. The yeast imparts the characteristic sherry flavor to the wine. No heating is used in this process.

Spoilage organisms. Wines, especially the sweet wines, are susceptible to spoilage by wild yeasts and bacteria. In the absence of air, spoilage is caused principally, by species of *Leuconostoc* and *Lactobacillus*. In the presence of air, species of *Acetobacter* may oxidize the alcohol of table wines to vinegar. Control measures are pasteurization of about 1 minute at 62.8°C, use of sulfur dioxide (SO_2), and sterilization by filtration through special filters. [F.M.M.; H.J.P.]

Bibliography: H. J. Grossman, *Grossman's Guide to Wines, Spirits and Beers*, rev. ed., 1955. L. A. Underkofler and R. J. Hickey (eds.), *Industrial fermentations*, vol. 1, 1954.

Wing

On an aircraft, a flat or slightly cambered plate, highly symmetric about a median plane (Fig. 1). The thickness of a wing is much smaller than its other dimensions (Fig. 2). A wing provides the lifting force that enables heavier-than-air craft to fly. Basically the action of a wing is similar to a venturi in that the length of air path along the top of the wing, greater than that along the bottom, causes the air to flow faster across the top, thereby producing an upward pressure differential. *See* BERNOULLI'S THEOREM; VENTURI TUBE.

Gross weight of the airplane divided by total wing area is a ratio called wing loading. To achieve low landing speeds, wings are lightly loaded; such wings, as in small private aircraft, may have external struts so that the structure can be light.

To avoid aerodynamic drag of the struts, espe-

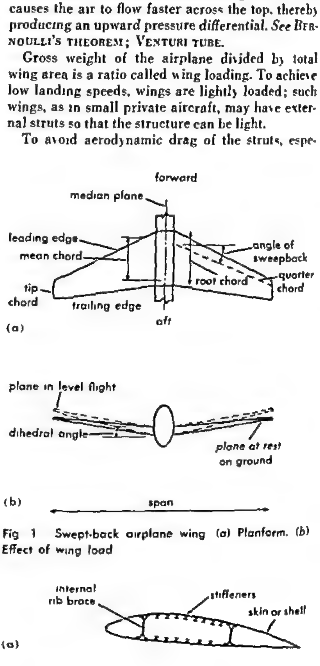


Fig. 1 Swept-back airplane wing. (a) Planform. (b) Effect of wing load.

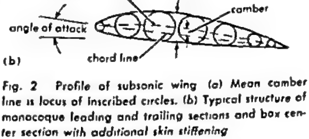


Fig. 2 Profile of subsonic wing. (a) Mean camber line is locus of inscribed circles. (b) Typical structure of monocoque leading and trailing sections and box center section with additional skin stiffening.

cally on high speed airplanes all structural members may be enclosed within the wing contour. The structure is then an internally braced cantilever. For torsional stiffness, the metal covering serves also as trusses above and below the internal braces, the arrangement being called a semimonocoque structure. In extremely thin wings where space prohibits internal braces, a full monocoque structure is used, the shell carrying the full load.

The wing need not be completely rigid. It may sag when at rest from its own weight or the weight of fuel tanks inside it and engines mounted in nacelles on it, and it may bow upward within its elastic limit from aerodynamic lift while in flight (Fig. 1). However, the aerodynamic properties of the wing should not change significantly in the presence of such deformations. *See* AEROELASTICITY.

When a swept back wing bends, its angle of attack in the streamwise direction is reduced (Fig. 3). For example, if flight conditions produce an upward force, a wing that is uniformly flexible from leading edge to trailing edge will deflect approximately the same distance along a line normal to its quarter chord. The deflection at the leading edge of a streamwise segment will then be less than at the trailing edge. The angle of attack along a streamwise segment will decrease with consequent decrease in lift. See AIRPLANE.

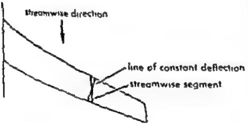


Fig. 3. Effect of bending on swept-back wing

The result of this behavior of a swept back wing is a tendency to shift the center of aerodynamic load inboard. Possibility of wing divergence is reduced, although the possibility of contact reversal is increased.

Engine nacelles are built into wings as part of the structure and as part of the aerodynamic form. However, the larger interruption of the fuselage has appreciable effect. With adequate fillets between wing halves and fuselage, the lift contributed by the fuselage can be almost equivalent to a central wing portion of like width. The drag of wing and fuselage together is considerably more than the sum of drags of the parts in isolation from each other. Wing tips also constitute a discontinuity that is not amenable to simple theoretical analysis. For these reasons exact performance is extrapolated from flight tests of similar vehicles or wind tunnel tests of scaled models.

Movable auxiliary fittings are added to wings so that the pilot can vary their characteristics in flight (Fig. 4). Lift increases at first with increasing angle of attack; however, at wide angles, the air separates from the top of the wing surface with abrupt decrease in lift, a condition called stall. A leading-edge slot that can be opened when greater-than-normal lift is needed admits air from the high-pressure region of the lower portion of the leading edge into the low-pressure region above the wing. *See* BOUNDARY-LAYER FLOW.

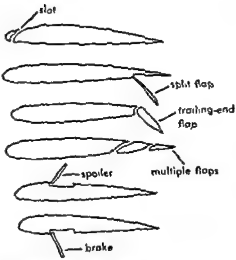


Fig. 4. Auxiliary wing fixtures for controlling characteristics.

Another means for increasing lift is to deflect a trailing edge flap into the air stream. The flap usually extends across the central portion of the wing leaving the outer portions of the trailing edge for ailerons (*see* FLIGHT CONTROLS). Both slots and flaps increase drag and so are suited only for use under such extreme conditions as take-off.

To limit the speed of an airplane in a dive or to decelerate it during landing, a spoiler can be deflected above the wing or a brake below it. The effects of these devices can be analyzed only approximately. Their exact characteristics are usually determined by experience and test. *See* AIRCRAFT TESTING; AIRFOIL, WING STRUCTURE.

[F H R]

Bibliography: R. L. Bisplinghoff, H. Ashley, and R. L. Halfman, *Aeroelasticity*, 1955; D. J. Peery, *Aircraft Structures*, 1950; R. von Mises, *Theory of Flight*, 1945.

Wing loading

The ratio of the weight W of an airplane to the area S of its wings. Wing loading W/S is expressed in lb/ft² or kg/m². Slow-speed airplanes have low wing loading. High-speed airplanes are heavily loaded, that is, their wing area is relatively small compared to their weight. Wing loading is a design criterion for both the airplane and the airfield. Propeller driven transports designed for minimum operating costs have the highest permissible wing loading. Landing-field distance is a direct function of wing loading, high wing loading requiring longer landing strips. *See* SAILPLANE; WING. [F H R]

tent and are taken to the winery in boxes or in bulk. Several pickings are required to produce fine wines. The grapes go through a crusher-stemmer, which crushes the berries, but not the seeds, and removes the stems. The must, still containing seeds and skins, is pumped to fermentation tanks where it receives about 100–150 ppm of sulfur dioxide (SO_2) gas. This controls, to a large extent, the growth of bacteria and wild yeasts, whereas the various strains of wine yeast are adapted to this amount of SO_2. A pure yeast starter, consisting of various strains of *Saccharomyces cerevisiae*, is then added. In the production of white wines, the juice is separated from the skins and seeds at an early stage of the fermentation, often after 4–8 hours. The residue may be pressed and extracted for the manufacture of lower quality wines or it may be fermented to completion and used for distillation. Many variations are in use in different wineries, depending on the type of wine to be made. The fermentation tanks, especially the large ones, are cooled to keep the temperature of the wine at 23.9–29.4°C.

After the initial fermentation is over, the wine is transferred to the storage cellar for completion of fermentation, clarification, aging, stabilization, and bottling. These operations are called cellar practices. The sediment of yeast and other insoluble matter is called lees. The new wine is drawn off, or racked, from the lees to avoid picking up undesirable flavors from the lees. Aging is then continued. Wines are often chilled in order to precipitate excess cream of tartar, the potassium acid tartrate from the grape, which otherwise might precipitate upon chilling of bottled wine. Champagne is made by allowing a secondary fermentation to occur with a special flocculating type of yeast, either in bottles, by the original French procedure, or in bulk, using pressure tanks followed by bottling. The CO_2 of sparkling wines is produced by yeast fermentation. Most sherry wines of California obtain their characteristic flavor and aroma by a heating process, called baking, over an extended period of time. Spanish sherries are made by the "flor" process, involving the participation of a film-forming yeast growing for many months at the surface of the wine in partially filled oak barrels. The [illegible] flavor to the [illegible]

[illegible] lly the sweet [illegible] wild yeasts and bacteria. In the absence of air, spoilage is caused principally, by species of *Leuconostoc* and *Lactobacillus*. In the presence of air, species of *Acetobacter* may oxidize the alcohol of table wines to vinegar. Control measures are pasteurization of about 1 minute at 62.8°C, use of sulfur dioxide (SO_2), and sterilization by filtration through special filters. [F.M.M., H.J.P.]

Bibliography: H. J. Grossman, *Grossman's Guide to Wines, Spirits and Beers*, rev. ed., 1955. L. A. Underkofler and R. J. Hickey (eds.), *Industrial Fermentations*, vol. 1, 1954.

Wing

On an aircraft, a flat or slightly cambered plate highly symmetric about a median plane (Fig. 1). The thickness of a wing is much smaller than its other dimensions (Fig. 2). A wing provides the lifting force that enables heavier-than-air craft to fly. Basically the action of a wing is similar to a venturi in that the length of air path along the top of the wing, greater than that along the bottom, causes the air to flow faster across the top, thereby producing an upward pressure differential. See BERNOULLI'S THEOREM; VENTURI TUBE.

Gross weight of the airplane divided by total wing area is a ratio called wing loading. To achieve low landing speeds, wings are lightly loaded; such wings, as in small private aircraft, may have external struts so that the structure can be light.

To avoid aerodynamic drag of the struts, espe-

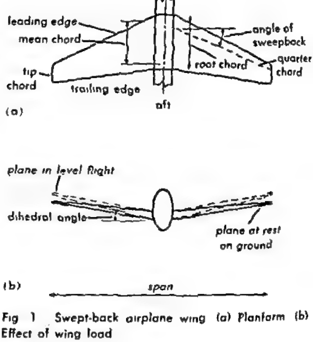


Fig 1 Swept-back airplane wing (a) Planform (b) Effect of wing load

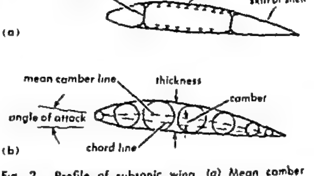


Fig 2. Profile of subsonic wing. (a) Mean camber line is locus of inscribed circles. (b) Typical structure of monocoque leading and trailing sections and box center section with additional skin stiffening

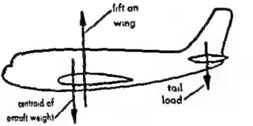


Fig. 2. Schematic diagram of major air loads

in service is much greater than this because the aircraft will encounter atmospheric gusts and because it is required to maneuver. The designed breaking strength of the structure is normally set at 1.5 times the maximum anticipated loads, the 1.5 multiplier being a factor of safety (*see* SAFETY FACTOR). The resulting load is called the ultimate design load

Maneuver loads. An aircraft must be capable of climbing, descending, turning, or, if it is a fighter, of executing rather violent acrobatics. Any such changes from a statically balanced condition produce load increments. These loads are generated when the pilot induces a sudden change in the angle of attack of the aircraft by manipulating the engine throttles and controls of his craft (*see* FLIGHT CONTROLS) An abrupt change in speed or angle of attack alters the lift coefficient or velocity and produces a change in total lift. The permissible degree of maneuverability of any specific model is based on its intended use The accompanying table presents general information on ranges of maximum anticipated load for various types of craft. Down increments of maneuver load are limited to 2 times the weight for transports and to somewhat higher values for the others The ultimate design load is 1.5 times the value shown

Maximum anticipated upward flight loads

Type of craft	Required lift force on craft
Transport	2.5 × weight
Bomber	2.5–4 × weight
Fighter	6–9 × weight
Special research	Up to 12 × weight

The multipliers in the table, which are called load factors, reflect the intended usage of the craft. Experience has shown these to be adequate, but the pilot must be aware of the maneuver limitations of his craft. It is possible for him to exceed these loads in flight. Figure 3 shows a simplified presentation of load factor and factor of safety applied to total external wing load. Tail load is assumed zero for simplicity of presentation. The presentation shows that the external air load equals the aircraft weight during cruise. In a 2.5*g* maneuver, maximum anticipated air load is 2.5 times the aircraft weight, the 2.5 being a design factor. On this basis, ultimate design load is 1.5 × 2.5 or 3.75 times the aircraft weight, the 1.5 being the safety factor.

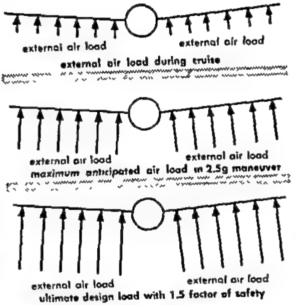


Fig 3 Total external air load.

Gust loads A second significant load produced on any craft's wings is an atmospheric gust. A high-velocity vertical gust produces an abrupt change in the direction of airflow over the wing. An upward gust produces an upward incremental force, and a down gust produces a downward incremental gust. The Civil Aeronautics Board (CAB) specifies the multiplying gust load factor.

For commercial aircraft, the normal gust velocity in ft/sec is specified as 30*K* at cruise speed, 15*K* at dive speed, and 40*K* for the speed recommended for flying under more turbulent conditions (less than cruise speed), where *K* is a coefficient dependent on wing loading, which is the ratio of aircraft weight to wing area. The recommended speed during turbulent conditions is determined from the aerodynamics of each aircraft type. The value *K* is about 1.20 for many typical current designs The factors must be derived for both upward and downward gusts, which produce loads on the wing comparable to the up and down maneuver loads The most critical among these loads must be determined for each element of the wing structure.

V-n diagram. The design speeds and load factors for any design are summarized in a graph called the *V-n* diagram (Fig. 4). The curve is derived from the specified requirements for the particular use of the vehicle. The curve applies to the craft as a whole. Balance considerations depicted in Fig. 2 permit the derivation of the total flight loads.

Landing and takeoff loads. Landing and takeoff loads are specified as part of the CAB air regulation for commercial vehicles and in connection with ground loads for military vehicles. Commercial aircraft must be capable of anticipated descent velocities of 10 ft/sec at design landing weight and 6 ft/sec at design takeoff weight. These descent velocities must be considered at different landing attitudes of the craft. Anticipated weight and

Wing structure

In an aircraft, the combination of outside fairing panels that provide the aerodynamic lifting surfaces and the inside supporting members that transmit the lifting force to the fuselage. The structure of a wing is an integration of the environment external to the vehicle wing, the aerodynamic shape of the wing, and the proposed use of the vehicle. The interaction of these three aspects of design leads to the selection of material, the general structural layout and, finally, to the detail choice of structural shapes, material thickness, joints, and attachments. The result is a structural framework covered with a metal skin that also contributes to the load-carrying function.

Structural materials. Wing structure has evolved from the early use of wood, doped canvas, and wire. The first general change was the replacement of wood with metal frameworks, although the wood and canvas structure is still found in lightweight personal aircraft. Doped canvas was replaced by light metal skins, which served only as fairing in the beginning but which were later designed to provide a portion of the structural strength. Today, aluminum alloy outer skins are prime structural elements on all commercial transports and on the great majority of military craft.

Magnesium, steel, and titanium are also used in internal primary structure and in local skin areas. The more exotic research craft, such as the X-15, are built at least partly of nickel alloy. Even less conventional materials may be used for tomorrow's reentry structures, which may or may not have wings (see REENTRY). Reinforced plastics, ceramics, and refractory metals are possibilities. A tentative design for the primary wing structure of a future reentry vehicle using advanced materials is shown in Fig. 1.

Design of a wing structure begins logically with a derivation of loads on the wing, both flight and net, as they are affected by particular design specifications; then the design proceeds through a choice of material to the final phases when the configurations of major structural units such as the prime wing box and the leading and trailing edge subassemblies are determined.

Total flight load. Flow of air past the wing's airfoil shape creates a lifting force (see AERODYNAMIC FORCE; AIRFOIL). Total lifting force L, in pounds, on an entire wing is

$$L = \rho V^2 C_L S/2$$

where ρ is air density in slugs, V is velocity in ft/sec, S is wing area in ft^2, and C_L is lift coefficient, which is a nondimensional coefficient derived from wind tunnel tests.

On a particular craft, the lifting force required of a wing depends on the craft's balance. For stability considerations, the aircraft is designed to maintain the centroid of any permissible loading condition forward of the center of lift on the total wing (Fig. 2).

The total load on the wing in normal cruising flight is the weight of the aircraft plus the balancing tail load. The total maximum anticipated load

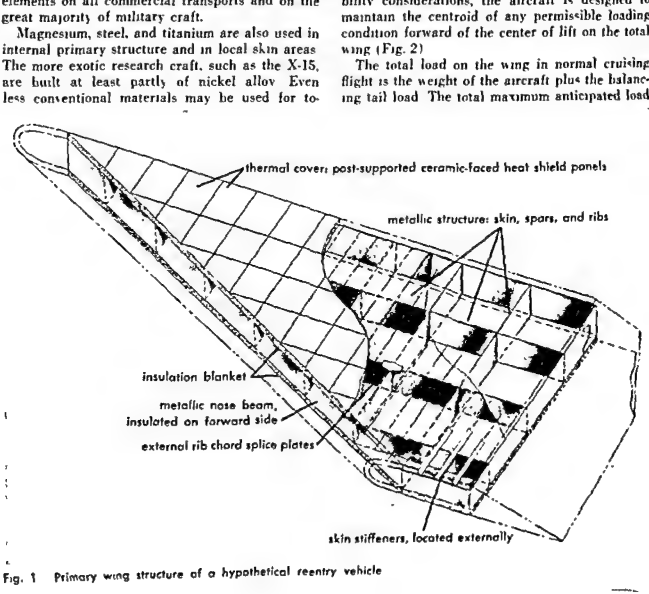


Fig. 1 Primary wing structure of a hypothetical reentry vehicle

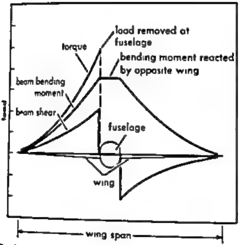


Fig 6 Typical shear, bending moment, and torque curves.

torsion forces produced by an eccentricity of the vertical and drag forces about the twisting center of the structure. These forces are presented as beam (vertical) and chord (drag) shear curves, beam and chord bending moments, and torque (Fig 6).

Choice of material. The derivation of the net I dc permits a quantitative consideration of the general structural framework Design of the structure, however, is inseparable from the choice of material The thickness and shape of the individual pieces also affect strength in lightweight, efficient structures because of compression stability modes.

High static strength and light weight—coupled with reasonable rigidity characteristics, good corrosion resistance, fatigue endurance, cost and formability—are prime considerations in choosing airframe material Highly loaded, efficient airframes built from 1930 to 1960 have found these qualities in aluminum. Aluminum has withstood the technical competition of steel, titanium, magnesium, wood, and reinforced plastics, although complete designs have been manufactured from each of these competitive materials to try them with complete realism. Recently aluminum has been somewhat supplanted by the requirements for high supersonic and hypersonic speeds in the atmosphere (the thermal thicket) and by the demands of space technology (the atmospheric reentry problem) But even in these problems, designers are considering protective shields and cooling of the structure to permit the use of aluminum with its favorable static strength, corrosion resistance, fatigue strength, cost, and formability, at so light a weight. It does appear, however, that the demands for speed interacting with the physical environment will find aluminum replaced by steel, refractories, ceramics, and composites of these A material's creep characteristics and strength at temperature are now important added parameters See AEROTHERMODYNAMICS

Authoritative data on strength properties of aircraft metals are available. Figures 7, 8, 9, and 10 present in graphical form comparisons of strength and stiffness with temperature. These curves indicate the superiority of aluminum at room temperature and disparity among materials as the temperature increases. For example, the ultimate tensile strength F_{tu} per unit weight is useful for application where tension loads predominate, such as on the wing lower surface (Fig. 7). The favorable tension ratios of titanium and molybdenum are negated by prohibitive cost or formability, as well as consideration of all properties shown in Figs 8, 9, and 10. The compressive yield strength (F_{cy}) per unit weight is useful where compressive stresses and compression stability considerations are prime factors, such as on the wing upper surface (Figs. 8 and 9). The plot of Young's modulus

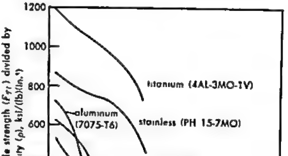


Fig 7 Ultimate tensile strength per unit weight for various materials

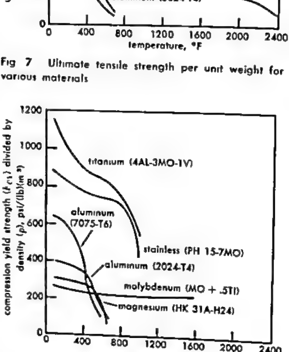


Fig 8 Compressive yield strength per unit weight for various materials.

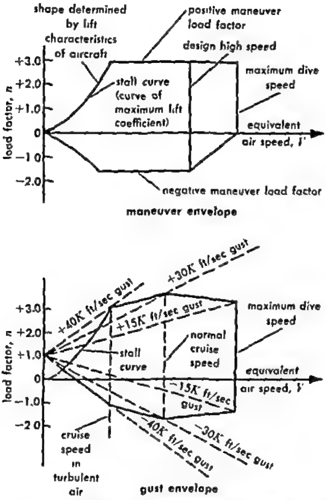


Fig. 4 *V-n diagrams of maximum anticipated flight loads*

centroids of design conditions are established by considering the vehicle usage. In addition, one-wheel landings, lateral drift, braking, turning, and pivoting are evaluated and defined in terms of loads. The dynamic effects of spinning the landing gear wheels at touchdown and the subsequent spring back of the gear and its attaching structure are critical conditions that are also evaluated. In general, the specified loadings and state-of-the-art development of landing gear shock-absorbing devices produce actual vertical loads on each main gear approximately equal to the aircraft weight (*see* LANDING GEAR). Maximum aft loads on the main gear are of the same order of magnitude. Side loads are 60–80% of this. This load envelope at landing usually determines the design of structure in the wing to which the main gear attaches. Because of the large twisting forces induced by the drag loads applied at ground level, which must be carried up and into the wing, this envelope is also critical for significant portions of the inboard wing structure. Ten or so landing conditions affect specific portions of the wing. The ultimate reaction of these loads is, of course, the weight and rotational inertia of the fuselage. The gear is expected to withstand actual anticipated loads without permanent distortion and ultimate design loads without structural failure.

Net wing loads. The maneuvers and gusts acting on the craft also generate accelerations and inertia effects. Therefore, the net loads on the wing, which the structure must sustain, are the net sum of the previously derived loads and the counteracting inertial effects. A simple example of a flight condition is presented in Fig. 5 for a hypothetical 100,000-lb aircraft where the tail load is assumed to be negligible. In essence, the wing is first assumed to be isolated from the rest of the craft. Then the air loads and inertia loads are applied, and the net of these is balanced with the remainder of the craft and applied to the wing through the wing-to-fuselage connections.

Fifteen to twenty-five different flight conditions are normally derived to satisfy possible critical conditions. These flight conditions encompass variations of aircraft gross weight and centroid location, fuel loadings, upward and downward maneuvers and gusts, special control surface and landing flap conditions, and any unusual dynamic loadings. The latter must consider dynamic interactions of unsteady airloads and structural stiffness (*see* AEROELASTICITY). Dynamic loads are also affected by concentrated weights such as engines, tip tanks, special pods, gun or rocket reactions.

Wing structure is examined at a number of wing cross sections in order to taper the weight of the material from the tip to the inner section of the wing. This is done by summing the net loads from the tip of the wing to the section in question. This calculation yields the vertical and drag shear forces, the bending forces produced by them as they are carried in toward the fuselage, and the twisting or

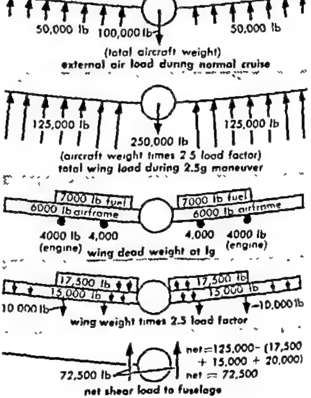


Fig. 5 Derivation of net loads

Fig 13. Partial tension field beam under test

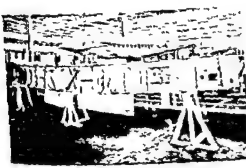


Fig. 14 Center section of Martin 404 wing beam

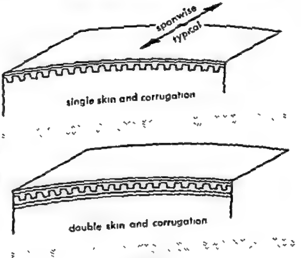


Fig 15 Types of top cover structure.

produce design shears in the rear webs In the inner portions of the wing, especially inboard of the landing gear, the landing conditions produce shears which exceed the flight loads by significant amounts These loadings require substantially thicker webs One may anticipate aluminum webs of 0 020-in thickness in the outer portions of the wing. These increase toward the wing root, and thicknesses of 0.125–0 188 in may occur at the inner section.

One of the beams from a twin-engine transport is shown in Fig 14 A portion of the beam at the centerline of the aircraft is of truss construction to permit access for equipment and for inspection and maintenance

Top of wing box The top of the prime box is the compression chord of the cantilever beam. It is also that portion of the box where the greatest variations in construction are found among the products of various manufacturers. The desirability of minimum weight and the requirements of the various positive and negative bending conditions have produced numerous configurations (Fig. 15).

These cover structures are designed as beam columns or plate structures (see BEAM COLUMN, PLATE, STRUCTURAL). They are subjected to axial load by the beam and chord bending moments, coupled with lateral loads from air pressures and, [illegible]

material sizes are established for a complete cross section, that is, for top of the wing box, bottom,

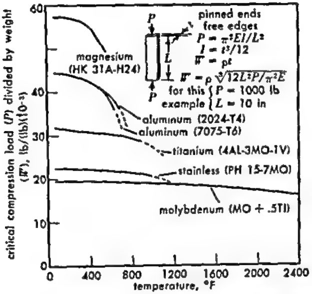


Fig 9. Critical compression loading in elastic region (Euler column) for various materials.

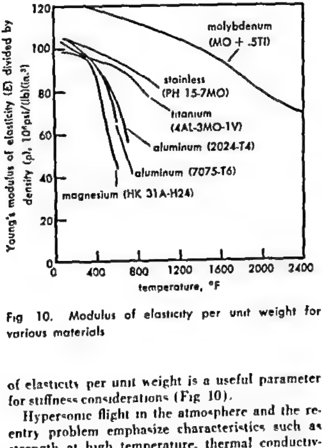


Fig 10. Modulus of elasticity per unit weight for various materials

of elasticity per unit weight is a useful parameter for stiffness considerations (Fig 10).

Hypersonic flight in the atmosphere and the reentry problem emphasize characteristics such as strength at high temperature, thermal conductivity, specific heat, coefficient of expansion, melting point, creep, and oxidation temperature, as well as the properties already mentioned. Data on the thermophysical properties for many of the solid materials that may be used in the future are being measured.

Prime structural framework. An airframe wing is essentially two cantilever beams joined together. Each wing tip is the free end of the cantilever, and the centerline of the vehicle represents the plane where the two fixed ends of the cantilevers are joined (Fig 11). The prime load-carrying portion of these cantilevers is a box beam made up usually of two or more vertical webs, plus a major portion of the upper and lower skins of the wing which serve as chords of the beam. This box section also provides torsional strength and rigidity (Fig. 12). Normally the prime box is designed to carry all the primary structural loads; these include all beam shears and bending moments, all drag shears and bending moments, and the torsional or twisting loads.

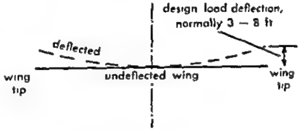


Fig 11. Wing tips deflect upward in normal flight.

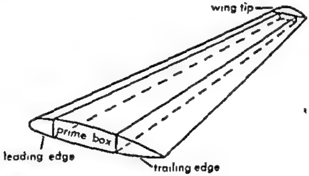


Fig 12 Prime box of wing half-span.

Leading and trailing edge portions of the wing, forward and aft of the prime box respectively, help to provide the airfoil shape required. These portions are designed to minimize their participation in the major load-carrying function. Where participation is forced by the detail design, the fasteners and materials reflect this, but normally the prime box strength is not reduced. An overlapping conservative assumption thus results.

Beam shear material The vertical webs of box beams are the prime load path for the beam or vertical shears These webs are comparable functionally to the webs of plate girders in highway or railroad bridges (*see* GIRDER, PLATE). The prime difference is the elastic buckling permitted in the aircraft web and consequently the much lighter material thickness This buckling is clearly shown in Fig 13, a photograph of a major aircraft beam under laboratory test where the shears imposed are buckling the web. When the load is removed, the buckles will disappear. This type of beam is called a partial tension field beam.

The web material of these beams is designed by net flight loads in the outer portion of the wing. The beam and torsional or twisting loads will normally combine to produce the loads which design the webs near the front of the wing. This is usually a large angle of attack condition. A low angle of attack condition, or one involving use of the control surfaces in the trailing edge region, will normally

e In this system, air that is heated by special ...ters or by the engine exhaust is passed along ... leading edge to melt the ice. This heated air ...mally exhausts at the wing tip.

Numerous aircraft have leading edge slats. These ...all airfoil sections are pulled out and forward ... and serve to main... ...ttack ...s the ...gle of attack is decreased these devices are withdrawn into the basic airfoil by springs (Fig. 18). ...lats are usually designed in short two-hinge ...ngths to eliminate wing deflection effects.

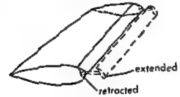


Fig 18 Leading edge slat

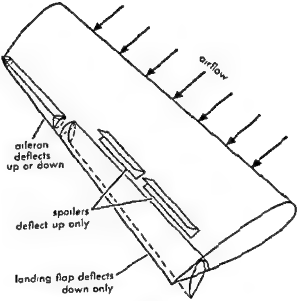


Fig 19. Trailing edge surfaces

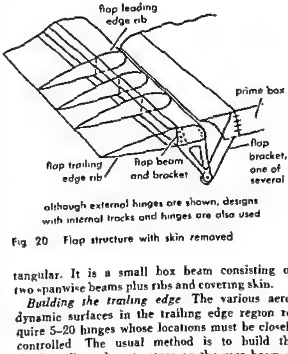
Trailing edge structures. The trailing edge structure is noteworthy for its various auxiliary devices which assist in aircraft control or in reducing aircraft landing speeds. Various forms of flaps, ailerons, and spoilers, with their hydraulic or electrical control mechanisms, fill the volume of space aft of the rear beam or spar (Fig. 19).

Landing flaps Landing flaps consist of a movable airfoil shaped structure located aft of the rear beam or spar They extend about two-thirds of the span of the wing Their aerodynamic function is to increase substantially the lift, thereby permitting lower takeoff and landing speeds

Flaps may be hung from the wing on two three, or four hinges. Structurally they represent a simple or continuous beam, depending on the number of hinges Where more than two hinges exist, the wing deflection effects must be considered in addition to local air loads. That is, as the wing deflects it will load the flap in bending it to the shape of the deflected wing. Flaps normally have a single spanwise main beam Leading edge ribs and trailing edge ribs are attached to this main beam and the airfoil shape is skinned over (Fig 20). Normal gages in these aluminum structures are between 0.020 and 0.040 in The hinge supports are cast or forged aluminum members The actual hinges may be plain or roller bearings

Ailerons. Ailerons are located near the tips of the wings in the trailing edge Their prime aerodynamic function is roll control of the aircraft. Structurally, the aileron is similar to the landing flap.

Spoilers. Spoilers are a special form of control surface and perform the aerodynamic function of the aileron, that is, roll control. They are located on the upper surface of the trailing edge at about midsemispan and deflect only upward by pilot control. The cross section of the spoiler is usually rectangular. It is a small box beam consisting of two spanwise beams plus ribs and covering skin.



although external hinges are shown, designs with internal tracks and hinges are also used

Fig 20 Flap structure with skin removed

Building the trailing edge The various aerodynamic surfaces in the trailing edge region require 5–20 hinges whose locations must be closely controlled The usual method is to build the fixed trailing edge structure on the rear beam or rear spar of the wing with a master tool holding the locations of the hinge points for the movable surfaces (Fig 21). The movable surfaces are then built from a matching hinge location tool Thus, when the assembled surface is brought to the wing it is certain to fit without binding

Wing-to-fuselage structure. The structural heart of the aircraft is the wing-to-fuselage joint (Fig 22) This connection is usually the most complex in form and in analysis Two major structural elements, the wing and the fuselage, with major loads running at right angles to one another, must be joined and analyzed for consistent deformations

calculated by methods fully described in texts. Repetitive refinement of material sizes finally results in an efficient top cover based on minimum weight and practical construction methods. Here, too, the gages begin with a minimum of 0.010-in. aluminum in honeycomb or 0.020-in. in corrugated forms. The plate designs begin with 0.050- to 0.060-in. minimum and reach thicknesses of 0.750 in. at inboard sections on some of the largest craft. The basic material thicknesses are determined for the compression loads and are virtually always adequate for load reversals which would put the top of the box in tension. Frequently, elements of the top of the box, including splices, are tested during the development of the craft to ensure adequate strength for the final design. *See* AIRCRAFT TESTING.

Bottom of the wing box. The bottom surface of the prime box is the tension chord of the beam. The bottom surface is most frequently a skin, tapering in thickness from a minimum at the wing tip to much heavier gages at the root. Specific gages are on the order of two-thirds to three-fourths of those for the top cover of the wing box because the allowable tensions are always higher than the permissible stresses for compression stability. Fatigue strength is a significant consideration in bottom surface design. Analyses and tests are used to provide the most reliable answers available. Spanwise reinforcing members on the bottom surface bring the compression strength up to requirements for load reversals. The lower surface must normally carry 35–40% of its tension allowable as a compression design load. Framing members around access openings help carry the stresses around such structural discontinuities in the lower wing frame (Fig. 16).

Wing ribs. At numerous places within the wing box, bulkhead-type structures called ribs are located. These internal structures serve to maintain the rectangular box shape and cut down the unsupported length of compression cover structures, to separate fuel tanks, and to distribute concentrated loads from guns, bombs, landing gear, or engines into the prime box. They are also located at any wing cross section where major load redistributions occur.

Wing weights and statistics. The weight of the aircraft divided by the wing area is called the wing load. From 1940 to 1960, wing loadings have steadily increased from 35 to 100 lb/ft² on modern high-speed aircraft.

While wing loading or intensity of loading has been increasing, the thickness of the wing has decreased from approximately 10% of chord (the distance from the leading edge to trailing edge) to 5% of chord. Thus, higher loadings on a thinner beam have increased the challenge to the designer to provide more strength in a shallower space. This has been done at the cost of increasing weights. In 1940, 4 lb/ft² represented a reasonable allowance for average wing weight. In 1960, a high-speed aircraft demands on the order of 10 lb/ft². The demands of hypersonic speeds and the resulting temperatures are not fully defined, but loadings above 10 lb/ft² may be anticipated.

Leading edge structure. The leading edge, or most forward portion of the wing, serves an important aerodynamic function in establishing smooth air flow and efficient lifting power for the wing (*see* AIRFOIL PROFILE). This area sustains the highest aerodynamic pressures. On hypersonic vehicles, it is also the hot spot of the design.

Structurally, the leading edge is an appendage, a fairing whose local loads must be supported by the prime wing box. This structure takes the form of cantilever beams or arch-type structure from the leading edge back to the front vertical beam (Fig. 17). These leading edge structures are normally of 0.032- to 0.072-in. aluminum, with skin covers in the same thickness range.

Frequently the leading edge of an aircraft incorporates anti-icing provisions. Two prime systems have been used in the past. One type is a rubber bladder stretched over the leading edge which is alternately inflated slightly and deflated. This action cracks any ice formation and the air stream sweeps it away. A second method is the hot leading

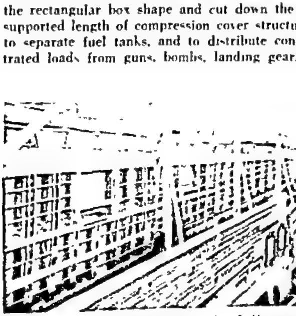


Fig. 16. Lower surface framework of Martin 404 outer wing

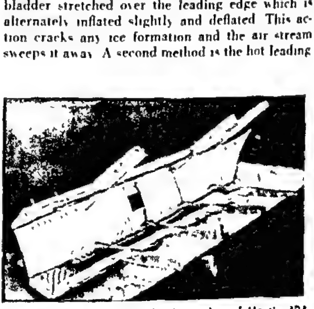


Fig. 17. Section of wing leading edge of Martin 404 transport

Wiring (electric)

A system of electric conductors, components, and apparatus for conveying electric power from source to the point of use. In general, electric wiring for light and power must convey energy safely and reliably with low power losses, and must deliver it to the point of use in adequate quantity at rated voltage. To accomplish this, many types of electric wiring systems and components are employed.

Electric wiring systems are designed to provide a practically constant voltage to the load within the capacity limits of the system. There are a few exceptions, notably series street-lighting circuits which operate at constant current.

In the United States the methods and materials used in the wiring of buildings are governed as to minimum requirements by the National Electric Code, municipal ordinances, and, in a few instances, state laws. The National Electrical Code is an American Standard approved by the American Standards Association. Most materials employed in wiring systems for light and power are tested and listed by Underwriters Laboratories, Inc. (UL). See ELECTRICAL CODES.

The building wiring system originates at a source of electric power, conventionally the distribution lines or network of an electric utility system. Power may also be supplied from a privately owned generating plant or, for emergency supply, a standby engine generator or battery.

The connection from the supply to the building system through the metering devices, main disconnecting means, and main overcurrent protection is the service entrance. The conductors, cables or busbars, are service conductors. The switch and fuse or circuit breaker serving as the disconnecting means and the main overcurrent protection is the service equipment. Up to six individual switches or circuit breakers may be used for the service entrance to a single building.

As a general rule only one set of service conductors to a building is permitted. Large industrial plants, commercial buildings, and institutions, however, are often served from more than one source. Separate service entrances are sometimes provided for emergency lighting, fire pumps, and similar loads.

Wiring systems. Wiring systems are generally three phase to conform to the supply systems. Energy is transformed to the desired voltage levels by a bank of three single-phase transformers. The transformers may be connected in either a delta or Y connection. With a delta connection the ends of the transformer windings are connected together and line conductors connected to these points. A three-phase, three-wire system is thus formed, from which a single phase line can be obtained from any two conductors. With the Y connection one end of each transformer winding is connected in a common connection, and line conductors are connected to the other ends of the transformer windings. This also forms a three-phase, three-wire system. A line wire is also often connected to the common point, forming a three-phase, four-wire system, from which a single-phase line may be obtained between the common wire and any other.

Service provided at the primary voltage of the utility distribution system, typically 13,800 or 4160 volts, is termed primary service. Service provided at secondary or utilization voltage, typically 120/208 or 277/480 volts, is called secondary service.

Primary service. Service at primary voltage levels is often provided for large industrial, commercial, and institutional buildings, where the higher voltage can be used to advantage for power distribution within the buildings.

Where primary service is provided, power is distributed at primary voltage from the main switchboard through feeders to load-center substations installed at appropriate locations throughout the building. The load-center substation consists of a high-voltage disconnect switch, a set of transformers, and a low-voltage switchboard enclosed in a heavy sheet-metal housing.

In practice, several feeder arrangements are employed. These include (1) single primary, a single

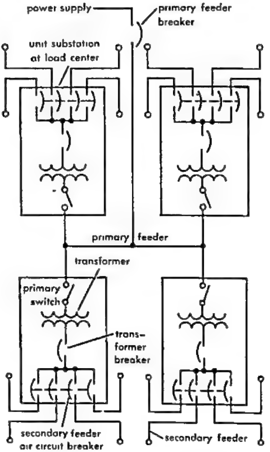


Fig. 1 Typical single primary distribution has a feeder from power supply to four load-center substations.

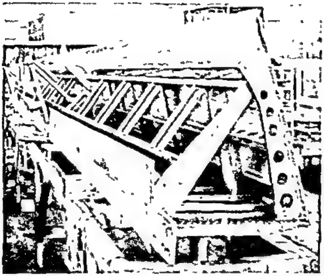


Fig 21. Trailing edge of wing with hinges for movable surfaces.

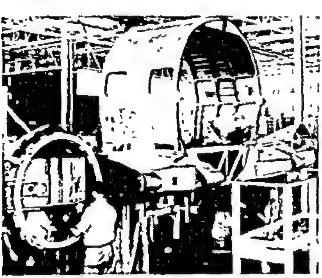


Fig 22. Center wing-to-fuselage connection.

The detail methods of indeterminate structural analyses are used The calculations may require anywhere from 10 to 100 redundants. High-speed digital computers are a necessity.

The wing structures which have been described are attached to heavy aluminum ring frames in the fuselage; these rings distribute the wing loads to the fuselage skins.

The assembly of the wing can be traced on the illustrations. Figure 14 is representative of a main spanwise beam. To this, leading edge structures, similar to that shown in Fig. 17, are added to produce a leading edge and front beam assembly In the same manner, a beam and the fixed trailing edge structures are joined together. Such an assembly is shown in Fig. 21. The rear beam and flap and aileron brackets are clearly indicated. These two major assemblies are then positioned in a major tool jig, and the intermediate beams, ribs, framing members, and top and bottom skins are progressively attached. Figure 16 shows such an assembly just before the attachment of the lower skins which close in the wing box. Figure 16 is an outer wing which eventually is attached as a unit to the center portion of the wing and fuselage. This center wing portion is shown in Fig. 22. The circular tubular structure at the left edge of Fig. 22 is part of the engine nacelle. *See* AIRFRAME. [A.J.K.]

Bibliography: Civil Aeronautics Board, *Civil Air Regulation*, pt 4b, amended to December 31, 1953, CAR4B, 1954; A. J. Kullas, *A Plan of Attack on Aircraft Fatigue*, SAE Preprint 386, 1954; W. M. Murray (ed.), *Fatigue and the Fracture of Metals*, 1952; L. E. Neville, *Aircraft Designers' Data Book*, 1950; A. S. Niles and J. S. Newell, *Airplane Structures*, vol. 1, 4th ed, 1954, vol. 2, 3d ed, 1943; S. P. Timoshenko, *Theory of Elastic Stability*, 1936; U.S. Dept. of the Air Force Air Research and Development Command, *Strength of Metal Aircraft Elements*, ANC-5, 1955; U.S. Munitions Board, Airways Committee, *Ground Loads*, ANC-2, 1952; WADC TR 58–476, *Thermophysical Properties of Solid Materials*, 1959

Wire

A thread or slender rod of metal. Wire is usually circular in cross section and is flexible If it is of such a diameter or composition that it is fairly stiff, it is termed rod. The wire may be of several small twisted or woven strands, but if used for lifting or in a structure, it is classed as cable Wire may be used structurally in tension, as in a suspension bridge, or as an electrical conductor, as in a power line. The working of metal into wire greatly increases its tensile strength. Thus, a cable of stranded small-diameter wires is stronger as well as more flexible than a corresponding solid rod. Wire may be treated or coated with various substances to protect it from corrosion or environmental influences In addition, electrical conducting wire is usually covered with insulating material. [F.H.R.]

Wire recording

Magnetic recording by use of a magnetized wire. It was first demonstrated by Valdemar Poulsen in 1898 Magnetic wire recorders were commercialized after World War II. However, since about 1950, the wire has been almost completely replaced by magnetic tape systems because tape offers many advantages

The wire used in magnetic wire recorders is a stainless steel alloy drawn to a diameter of 0.004 in. Smaller and larger diameters have also been used. Except for minor details, the techniques and systems used for magnetic wire recording are similar to those for magnetic tape For example, in the magnetic head, a slot is provided for the wire so that a more uniform magnetization of the wire is provided The usual recording speed of the wire is 24 inches per second (ips). However, other speeds have also been used The wire is driven at a constant speed by a grooved capstan and is wound on a spool. At a speed of 24 ips, a 1-hour program can be stored on a spool 2¾ in. in diameter and ⅝ in in width. *See* DISK RECORDING; MAGNETIC RECORDING. [H.E.O.]

Wiring methods. Methods of wiring in common use for light and power circuits are as follows: (1) insulated wires and cables in raceways; (2) nonmetallic sheathed cables; (3) metallic armored cables; (4) busways; (5) copper-jacketed, mineral-insulated cables; (6) open wiring on insulations; (7) nonmetallic sheathed and armored cables in cable troughs; and (8) nonmetallic waterproof wiring.

Raceways in which insulated conductors may be installed are of several types: (1) ...

metal conduit, (5) surface metal raceway; (6) underfloor raceway, (7) cellular floor raceway; and (8) wireway.

The selection of the wiring method or methods is governed by a variety of considerations, which usually include code rules limiting the use of certain types of wiring materials; suitability for structural and environmental conditions; installation (exposed or concealed); accessibility for changes and alterations; and costs. Several methods may be employed together, for example, feeder busway risers in a multistory office building with the rest of the wiring in rigid conduit and ...

Circuit de... system is ... loads, establishing the branch-circuit and feeder requirements, and then determining the service-entrance requirements. Outlets for lighting fixtures, motors, portable appliances, and other utilization devices are indicated on the building plans and the load requirement of each outlet noted in watts or horsepower. Lighting fixtures and plug receptacles are then grouped on branch ... to the ...

Lighting ... ated in the approximate center of the loads ... serve; however, other considerations may require other locations. Panelboards in commercial and institutional buildings usually are located in corridors or electric closets. The size and number of panelboards is determined by the number of branch circuits to be served.

Lighting loads. Lighting branch circuits may be loaded to 80% of circuit capacity. However, there is a reasonable probability that the lighting equipment will be replaced at some future time by equipment of higher output and greater load. Therefore, in modern practice, lighting branch circuits are loaded only to about 50% capacity, typically not more than 1200 watts on a 20-ampere (amp) branch circuit.

Lighting branch circuits are usually rated at 20 amp. Smaller 15-amp branch circuits are used mostly in residences. Larger 30- and 50-amp branch circuits are limited to installation of large lamps with heavy-duty lamp holders approved for connection to such circuits. The minimum conductor size for 20-amp branch circuits is No. 12; however, on long runs, larger conductors may be required to avoid excessive voltage drop. A common practice is to use No. 12 conductors between outlets and No. 10 conductors for the connection between the first outlet and the lighting panelboard.

Motor loads. These power loads are usually served by individual branch circuits, which must be rated not less than 125% of the full-load rating of the motor.

Feeder and service-entrance design. The sum of the branch-circuit loads, including future load allowances, determines the feeder load. The National Electrical Code provides a table of factors for various occupancies giving the minimum unit load per square foot and demand factors that may be applied. Demand factors are applied in installations where the loads are diversified and not likely to occur at one time. The number of feeders and their loads determines the number and size of distribution panelboard circuit elements required. The sum of the feeder loads determines the size and capacity of the service-entrance conductors and equipment.

Conductor sizes. The size of wires and cables used in electrical wiring systems is expressed in terms of the American Wire Gage (AWG), known also as the Brown and Sharpe (B&S) gage. Size designations run from No. 14, the smallest size commonly used in wiring systems for light and power, to No. 4/0, the largest size in the gage. Sizes larger than No. 4/0 are designated by their cross-section areas expressed in circular mils. The largest size in practical usage is 2,000,000 circular mils. A circular mil is the area of a circle 0.001 in. in diameter.

Conductor capacity. The current-carrying capacity of wiring conductors is determined by the maximum insulation temperature that can be tolerated and the rate at which heat can be dissipated. All conductors offer some resistance to the flow of electric current. Consequently, heat is produced in the conductor by the flow of current through its resistance. The amount of heat is determined by the square of the current in amperes times the resistance in ohms (I^2R). Conductor heat is dispersed through the insulation and the surrounding raceway to the air. See INSULATION, ELECTRIC.

In practice, maximum current-carrying capacity of conductors is set forth in standard tables developed from laboratory tests and field experience. The National Electrical Code specifies the maximum current-carrying capacity of conductors. For any given size of conductor, the maximum capacity varies with the type of installation (exposed or in raceways) and the maximum safe temperature of the insulation. Approved values are reduced for high ambient temperatures and for more than three current-carrying conductors in a single raceway. See CONDUCTOR, ELECTRIC.

Feeders supplying several motors must be rated at not less than 125% of the full-load current of the largest motor plus 100% of the full-load currents of the remaining motors.

Feeders serving continuous loads that are likely to operate for long periods of time (as office ...

primary feeder serving several substations; (2) multi-primary, individual primary feeders to each substation; (3) loop primary, two primary feeders serving several substations, interconnected to form a ring or loop; and (4) primary selective, two primary feeders serving several substations, connectable to either feeder by means of a selector switch.

Wiring from the low-voltage-switchboard end of the substation to the load follows conventional utilization-voltage practice.

In large industrial plants, the secondary circuits of several substations are sometimes interconnected by feeder ties through switches or circuit breakers. The switches may be closed to form a network, provided that the transformers are suitable for parallel operation, or operated to transfer the load from one substation to another.

Enclosed bus-bar systems, called busways, are frequently used in the wiring system of industrial plants and large buildings. Busways are made and shipped in standard lengths with a wide variety of fittings They are connected together on the job and installed as service-entrance conductors or feeders. One type of busway is designed to receive bus plugs at intervals along its length and thus functions both as a feeder and as a panelboard The plugs which tap the feeder may contain fuses or circuit breakers, but the tap is often carried down to a more readily accessible overcurrent protective device.

Secondary service. This type of service supplies power to the building at utilization voltage. Most secondary services in the United States are 120/208 volts, three-phase, four-wire, or 120/240 volts single-phase, three-wire serving both light and power In some communities, separate light and power services are provided, typically 120/240 volts single-phase for lighting and 440 volts, three-phase, delta for power.

Three-phase, four-wire services are almost universally Y with the neutral at ground potential In some instances, however, a three-phase, four-wire service may be delta, with the neutral connected at the center tap of one phase. Such services provide typically, 240 volts, three-phase, three-wire for power and 120/240 volts, single-phase, three-wire for lighting.

For relatively large buildings where the loads are predominantly fluorescent lighting and power (as for air conditioning), the service is often 277/480 volts, three-phase, four-wire, supplying 480 volts for power and 277 volts, phase-to-neutral, for the lighting fixtures.

Distribution switchboards. From the service entrance, power is carried in feeders to the main switchboard, then to distribution panelboards. Smaller feeders extend from the distribution panelboards to light and power panelboards. Branch circuits then carry power to the outlets serving the various lighting fixtures, plug receptacles, motors, or other utilization devices.

The main distribution switchboard may also include the service equipment in its assembly. It consists of a group of switches and fuses or circuit breakers in a sheet-metal enclosure. It provides individual disconnecting and overcurrent protection for each feeder. In large buildings, additional distribution panelboards may be located at load centers

Light and power panelboards provide individual disconnecting and overcurrent protection for the branch circuits The circuit breakers of lighting panelboards are sometimes used as switches to operate the lighting circuits.

Plug-receptacle power at 120/240 volts, such as power for small appliances and business machines, in buildings provided with 277/480-volt supply is obtained from transformers A feeder circuit from a 277/480-volt panelboard energizes the transformer primary. The secondary feeder serves a separate panelboard. The branch circuits serving the plug-receptacle outlets are conventional.

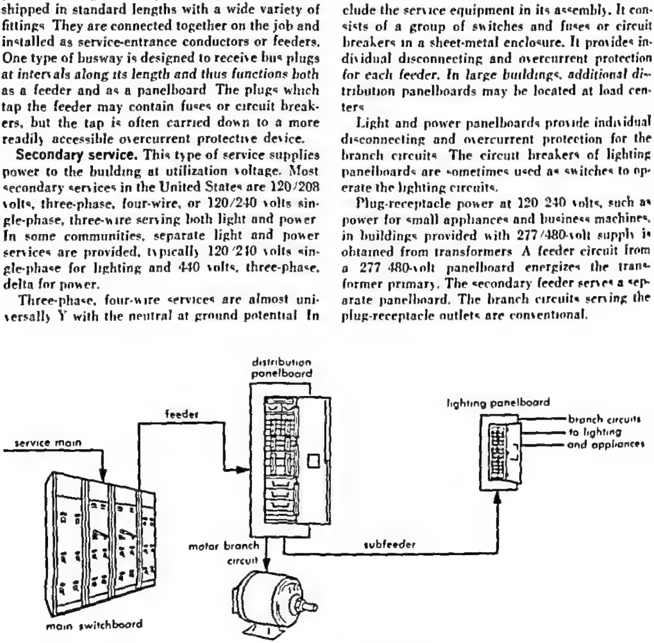
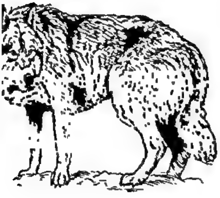


Fig 2 Distribution wiring system has components to supply utilization voltage for several types of load



The gray wolf, *Canis nubilus*, length male 64 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill 1949)

Wolframite

A mineral with chemical composition $(Fe,Mn)WO_4$, intermediate between ferberite, the iron tungstate, and huebnerite, the manganese tungstate, which form a complete solid solution series. *See* FERBERITE, HUEBNERITE.

Wolframite occurs commonly in short, brownish-black, monoclinic, prismatic, bladed crystals. It is difficult to dissolve in acids, but wolframite high in iron fuses readily to a magnetic globule.

Wolframite is probably the most important tungsten mineral. A quick and easy test for tungsten is to fuse the mineral powder with charcoal and sodium carbonate and boil the residue in hydrochloric acid with a few grains of granulated metallic tin. The presence of tungsten gives the solution a Prussian-blue color. The major use of tungsten is in making ferrous (steel) and nonferrous alloys, tungsten carbide, metallic tungsten, and tungsten chemicals. *See* TUNGSTEN.

Wolframite is found associated with quartz in veins in the peripheral areas of granitic bodies. It is also found in veins associated with sulfide minerals such as pyrite, chalcopyrite, arsenopyrite, bismuthinite, together with cassiterite, molybdenite, hematite, magnetite, tourmaline, and apatite. It also occurs in placers.

China is the major producer of wolframite. Tungsten minerals of the wolframite series occur in many areas of the western United States, of which the major producing district is Boulder County and northern Gilpin County in Colorado.

[E.C.T.C.]

Wolf-Rayet star

A member of a class of very hot stars (100,000–50,000°K) which characteristically show broad, bright emission lines in their spectra. Wolf-Rayet stars are classified into two main groups, the WN (nitrogen) and the WC (carbon) stars. The emission lines show very high excitation (lines of He II and even up to O VI are strong) and are broadened by an as yet unexplained mechanism. Displaced absorption lines caused by fast-moving gas shells are sometimes seen. The simple explanation of expansion of an outer layer with velocities up to 2000 km/sec is found insufficient from studies of Wolf-Rayet stars which happen to be members of a spectroscopic binary system. Scattering of light by fast-moving electrons has also been suggested. Luminosities must be very high, in the range of 10^4–10^5 times that of the Sun. These stars are probably very young and represent an unstable and short-lived early stage in stellar evolution. *See* STAR; STELLAR EVOLUTION.

[J.L.GR.]

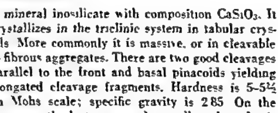
Wollastonite

A mineral inosilicate with composition $CaSiO_3$. It crystallizes in the triclinic system in tabular crystals. More commonly it is massive, or in cleavable to fibrous aggregates. There are two good cleavages parallel to the front and basal pinacoids yielding elongated cleavage fragments. Hardness is 5–5½ on Mohs scale; specific gravity is 2.85. On the cleavages the luster is pearly or silky, the color is white to gray. Wollastonite occurs characteristically as a contact metamorphic mineral in limestones where it is associated with diopside, garnet, tremolite, idocrase, and epidote. It is found in large masses in the Black Forest of Germany; Brittany, France; Chiapas, Mexico; and Willsboro, New York, where it is mined as a ceramic material. *See* SILICATE MINERALS.

[C.S.HU.]

Wolverine

A large carnivorous mammal, *Gulo luscus*, of the family Mustelidae, found chiefly in the Hudson Bay area. The wolverine is about 3 ft long, with a short, shaggy tail, and long blackish-brown fur. It has a broad yellow band on the side. The fur is valuable because it does not collect moisture, thus making it ideal to form the face strip on parkas. The wolverine is a serious enemy of northern trappers, wrecking trap lines and food caches with equal thoroughness. It is a wily, powerful, and vicious animal, capable of killing caribou and bear. Although never abundant, the wolverine was once far more common than at present. *See* CARNIVORA.



The wolverine, *Gulo luscus*, length 3 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

[J.D.B.]

building lighting) should not be loaded to more than 80% of rated capacity. The size of the feeder conductors will often be determined by the permissible voltage drop, which may require larger conductors than would be required by current-carrying capacity considerations alone.

Voltage drop. A drop in voltage along a conductor is a characteristic of all electric circuits. The voltage drop in a circuit causes the voltage at the load to be less than the voltage applied to the circuit.

Wire or cable circuits are designed to carry a certain load but, whether the conductor is copper, aluminum, or other metal, the resistance characteristic impedes the flow of current.

Since conductor resistance is proportional to conductor length, longer circuits are especially susceptible to excessive voltage drops. The amount of voltage drop E_D is determined by the current I in amperes times the resistance R in ohms ($E_D = IR$).

Good practice in circuit design dictates the following percentage values for maximum voltage drop: 5% from service entrance to any motor load; 2% from service entrance to any panelboard; 2% from panelboard to any outlet on branch circuit; 4% in feeders and 1% in branch circuit to motor; and 2% total in conductors to electric heating equipment.

The feeder voltage drop is calculated by formulas derived from Ohm's law and the resistance of the conductor. For three-wire, three-phase circuits (neglecting inductance), the voltage drop E_D is:

$$E_D = \frac{1.732KIL}{CM}$$

in which K is the resistance of a circular mil-foot of wire (for copper, 10.8 ohms), I is the current in amperes, L is the circuit length (source to load) in feet, and CM is the cross-section area of the conductor in circular mils

Copper loss This characteristic of a circuit is related to voltage drop. It is a power loss, designated as the I^2R loss, and is often expressed in percentage as the ratio of the wattage loss to the wattage delivered to the circuit.

Circuit protection. In wiring systems of high capacity (typically 1200 amp and above) supplied from utility networks of large capability, overcurrent protective devices of high interrupting capacity are required. Circuit breakers of special design or current-limiting fuses are employed in such installations. In some cases, current-limiting reactors or bus-bar arrangements presenting appreciable reactance under short-circuit conditions are inserted in the service conductors. *See* ELECTRIC PROTECTIVE DEVICES [W.T.S.]

Bibliography: A. L. Abbott, *National Electrical Code Handbook*, 9th ed., 1958; A. E. Knowlton, *Standard Handbook for Electrical Engineers*, 9th ed., 1957; J. F. McPartland et al., *Electrical System Design*, 1956

Wiring diagram

A drawing that illustrates both the electrical and mechanical relationship between parts on a component between which electrical wiring must be connected. A wiring diagram is distinguished from an electrical schematic in that the arrangement of the schematic bears no necessary relationship to the mechanical arrangement of the electrical elements in the component. The wiring diagram provides an accurate picture of how the wiring on the components and between components should appear in order that the electrical wiring technician can install the wiring in the manner that will best contribute to the optimum performance of the device. The degree of symbolism used in a wiring diagram depends on the extent of standardization in the particular field. For example, in telephone switchboard wiring, which consists of many standardized repetitive operations, extensive symbolism is used. On the other hand, when the exact physical location of wiring is important, as in radio-frequency devices where electromagnetic and electrostatic coupling between wires is appreciable, the diagram can be quite pictorial.

Wiring diagrams also include such information as type of wire, color coding, methods of wire termination, and methods of wire and cable clamping. *See* SCHEMATIC DRAWING. [R.W.M.]

Witherite

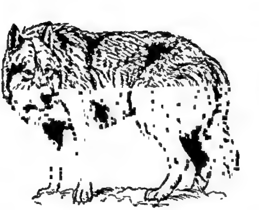
The mineral form of barium carbonate. Witherite has orthorhombic symmetry and the aragonite structure-type. Crystals, often twinned, may appear hexagonal in outline. It may be white or gray with yellow, brown, or green tints. Its hardness is 3½ and its specific gravity 4.3

Witherite may be found in veins with barite and galena. It is found in many places in Europe, and large crystals occur at Rosiclare, Illinois. *See* BARIUM; CARBONATE MINERALS. [R.I.H.]

Wolf

Any of several large, intelligent, nocturnal members of the family Canidae. The gray, or timber wolf, *Canis nubilus*, formerly was common in the forested areas of most of the Northern Hemisphere but is now extinct over much of its former range and rare over most of the remainder. It is common today only in the more remote parts of North America and Siberia. It is the largest member of the family, reaching a length in excess of 5 ft. The color of individuals and races varies widely, from almost black to virtually white in the Arctic form. Wolves hunt in packs and are effective predators, especially on deer and other large game mammals. they also eat birds, rabbits and other small animals, carrion, fruits, and berries. (*See* illustration on following page.)

Canis niger, the red wolf, is a smaller form, now limited to eastern Texas and the Ouachita Ozarks upland area. *See* CARNIVORA [J.D.B.]



The gray wolf, *Canis nubilus*, length male 64 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Wolframite

A mineral with chemical composition $(Fe,Mn)WO_4$, intermediate between ferberite, the iron tungstate, and huebnerite, the manganese tungstate, which form a complete solid solution series. *See* FERBERITE; HUEBNERITE

Wolframite occurs commonly in short, brownish-black, monoclinic, prismatic, bladed crystals. It is difficult to dissol[illegible] ed mineral. A quick and easy test for tungsten is to fuse the mineral powder with charcoal and sodium carbonate and boil the residue in hydrochloric acid with a few grains of granulated metallic tin. The presence of tungsten gives the solution a Prussian blue color. The major use of tungsten is in making ferrous (steel) and nonferrous alloys, tungsten carbide, metallic tungsten, and tungsten chemicals. *See* TUNGSTEN.

Wolframite is found associated with quartz in veins in the peripheral areas of granitic bodies. It is also found in veins associated with sulfide minerals such as pyrite, chalcopyrite, arsenopyrite, bismuthinite, together with cassiterite, molybdenite, hematite, magnetite, tourmaline, and apatite. It also occurs in placers.

China is the major producer of wolframite. Tungsten minerals of the wolframite series occur in many areas of the western United States, of which the major producing district is Boulder County and northern Gilpin County in Colorado. [E.C.T.C.]

Wolf-Rayet star

A member of a class of very hot stars (100,000–35,000°K) which characteristically show broad, bright emission lines in their spectra. Wolf-Rayet stars are classified into two main groups, the WN (nitrogen) and the WC (carbon) stars. The emission lines show very high excitation (lines of He^{+} and even up to O[illegible] are strong) and are broadened by an as yet unexplained mechanism. Displaced absorption lines caused by fast-moving gas shells are sometimes seen. The simple explanation of expansion of an outer layer with velocities up to 2000 km/sec is found insufficient from studies of Wolf-Rayet stars which happen to be members of a spectroscopic binary system. Scattering of light by fast-moving electrons has also been suggested. Luminosities must be very high, in the range of 10^4–10^5 times that of the Sun. These stars are probably very young and represent an unstable and short-lived early stage in stellar evolution. *See* STAR; STELLAR EVOLUTION. [J.L.GR.]

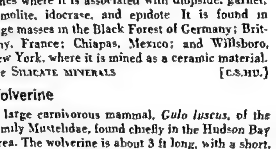
Wollastonite

A mineral inosilicate with composition $CaSiO_3$. It crystallizes in the triclinic system in tabular crystals. More commonly it is massive, or in cleavable to fibrous aggregates. There are two good cleavages parallel to the front and basal pinacoids yielding elongated cleavage fragments. Hardness is 5–5½ on Mohs scale; specific gravity is 2.85. On the cleavages the luster is pearly or silky; the color is white to gray. Wollastonite occurs characteristically as a contact metamorphic mineral in limestones where it is associated with diopside, garnet, tremolite, idocrase, and epidote. It is found in large masses in the Black Forest of Germany; Brittany, France; Chiapas, Mexico; and Willsboro, New York, where it is mined as a ceramic material. See SILICATE MINERALS [C.S.HU.]

Wolverine

A large carnivorous mammal, *Gulo luscus*, of the family Mustelidae, found chiefly in the Hudson Bay area. The wolverine is about 3 ft long, with a short, shaggy tail, and long blackish brown fur. It has a broad yellow band on the side. The fur is valuable because it does not collect moisture, thus making it ideal to form the face strip on parkas. The wolverine is a serious enemy of northern trappers, wrecking trap lines and food caches with equal thoroughness. It is a wily, powerful, and vicious animal, capable of killing caribou and bear. Although never abundant, the wolverine was once far more common than at present. *See* CARNIVORA

[J.D.B.]



The wolverine, *Gulo luscus*, length 3 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Wood (anatomy and identification)

The study of the structure, especially the microscopic structure, of wood, or xylem (*see* XYLEM).

Gross features of wood. A cross section of a tree grown in a temperate region reveals the central, darker, dead heartwood, and the outer, lighter, living (at least in part) sapwood. Each of these regions consists of concentric bands, or growth rings. Usually one growth ring is formed during a growing season; however, some tropical trees lack

and the cell walls usually thinner, and the late wood (called *summer wood in temperate-zone trees*), in which the cells are smaller and the cell walls are usually thicker. *See* STEM (BOTANY).

Anatomy of softwoods. The structure of a softwood or gymnospermous wood, such as pine, is relatively simple and homogeneous (*see* GYMNOSPERMAE). Most of the wood consists of long, empty, spindle-shaped cells called tracheids, which bear bordered pits in their side walls (Fig. 1). There are also thin bands of bricklike, living cells radiating from the center. These vascular rays are uniseriate, or only one cell wide as seen in cross section. In addition, some woods of gymnosperms have resin canals (canals lined with secretory parenchyma cells), and some gymnosperms have axial or vertical wood parenchyma cells. *See* PARENCHYMA; SECRETORY STRUCTURES, PLANT.

Anatomy of hardwoods. The structure of a hardwood or angiospermous wood, such as oak, is much more complex and heterogeneous than that of a gymnospermous wood (*see* ANGIOSPERMAE). In addition to tracheids and uniseriate vascular rays there are vessel elements, wood fibers, axial wood parenchyma cells, and multiseriate vascular rays.

Vessel elements are dead conducting cells, resembling segments of tile, that fit together to form tubes (vessels). The end walls (plates) of vessel elements may have several openings or perforations (scalariform perforation plates), or one perforation (simple perforation plates). If the vessels, as seen in cross section, are larger or more numerous in the early wood than in the late wood, the wood is said to be ring-porous; if the vessels are evenly distributed and of uniform size, the wood is classified as a diffuse-porous wood (Fig. 2). As seen in cross section, vessels (pores) are termed solitary pores, if they occur singly; pore multiples, if they occur in radial groups of two or more vessels that are crowded and flattened together; pore clusters, if the groupings are irregular; and pore chains, if the pores form a radial series of isolated pores.

The wood fibers may be tracheids with large bordered pits, or fiber-tracheids with bordered pits smaller than those in adjacent vessel elements, or libriform wood fibers with simple pits. The last two elements are dead cells with thick walls that give support. The presence of these thick-walled fibers

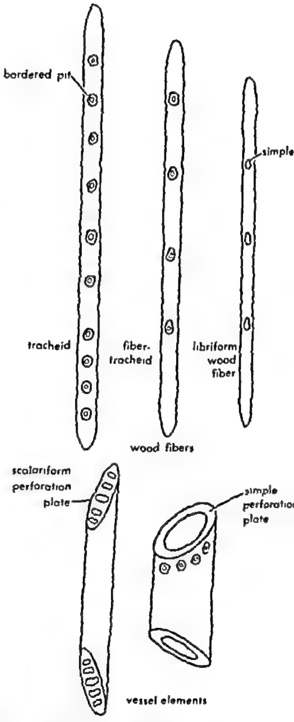


Fig. 1 Wood fibers and vessel elements

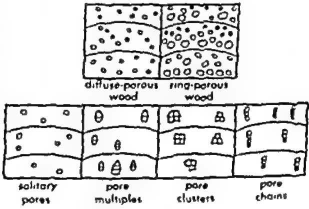


Fig. 2 Pore distribution in wood

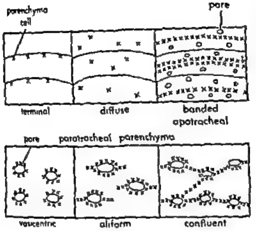


Fig 3 Wood parenchyma distribution

in angiosperm xylem is the reason why these woods are called hardwoods.

The axial or longitudinal wood parenchyma cells are living bricklike storage cells that are arranged in vertical or longitudinal series (Fig 3). As seen in cross section wood parenchyma [illegible]

scattered throughout the cross section), banded apotracheal (bands of parenchyma occurring independently of the pores), and paratracheal (parenchyma associated with vessels). In the latter category, the parenchyma cells may be classified as vasicentric (forming a complete sheath around vessels), aliform (vasicentric with winglike extensions), or confluent (aliform parenchyma whose extensions join to form irregular tangential or diagonal bands).

Angiosperm woods commonly have both uniseriate and multiseriate (two or more cells wide) vascular rays If the rays are composed only of radially elongated cells, they are called homogeneous rays, if composed of radially elongated cells and vertically elongated or square cells, they are termed heterogeneous rays (Fig 4)

Identification of wood. Woods are identified by utilizing the features described above, because the various genera and often the species, possess unique combinations of these characteristics In practice a key is employed which consists of groups of paired, contrasting statements. By reading these and progressively selecting those characteristics that apply to the wood being identified, the name of the wood is found The detailed description of the wood and the photomicrographs are checked to see

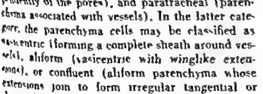


Fig 4 Vascular rays

if the unknown wood corresponds to the description and the photographs. Usually the key begins by separating the woods without vessels (gymnosperms) from those with vessels (angiosperms). Under the latter, the ring-porous woods are separated from the diffuse-porous woods. Then by referring to such features as vessel-element type, pore distribution, parenchyma distribution, type of fiber, and type of ray, the proper species name is found.

Phylogenetic relationships. Plant anatomists have determined the phylogenetic or evolutionary development of many of these cells. *See* PLANT EVOLUTION. The tracheid is obviously the primitive or basic cell in the xylem During the course of the evolutionary history of vascular plants, tracheids have been modified to form progressively more efficient mechanical elements (fiber-tracheids and then libriform wood fibers). In another direction, involving improvement in conduction, tracheids have gradually been replaced by vessel elements with scalariform perforation plates and then by vessel elements with simple perforation plates Heterogeneous rays are considered more primitive than homogeneous rays.

By means of these lines of evolutionary specialization, the wood anatomist is able to make contributions to plant phylogeny, the science which endeavors to unravel the relationships among plant groups. Thus, it has been demonstrated fairly conclusively that the Amentiferae (oaks, beeches, elms) are not primitive angiosperms, as was once supposed, they are much more advanced than the Ranales (magnolias). Wood anatomy has likewise assisted in the phylogenetic placement of many other groups.

Cell walls. Most wood cells possess three cell walls: the intercellular layer, or middle lamella, which consists of pectic substances and *lignin*, the primary wall, which is composed of pectic substances, cellulose, and lignin; and the secondary wall, which consists of cellulose and lignin. *See* CELL WALLS IN PLANTS, CELLULOSE, PECTIN

Microtechnique. To prepare woody tissue for microscopic examination, woods may be taken from living trees in which case the tissues are killed in a fixative such as chromoacetic acid, or they may be cut from dead timbers (*see* MICROTECHNIQUE). In the latter case they are boiled in water to remove air, or the air may be removed by means of a pump Usually both the wood from living trees and that from dead timbers is treated with hydrofluoric acid to delignify and soften the wood Thin cross, radial, and tangential sections are then cut with a sliding microtome The sections are stained with dyes such as hematoxylin and safranin The former stains the middle lamella (pectin) blue and the latter stains the secondary walls (*lignin*) red. The stained sections are then mounted in balsam between a glass slide and a cover glass After a drying period, the slides are ready to be studied with the microscope (*see* MICROSCOPE, OPTICAL)

If individual, whole-wood cells are to be studied,

the wood cells are dissociated by dissolving the middle lamellae by means of treatment with macerating fluids, such as equal parts of 10% chromic acid and 10% nitric acid.

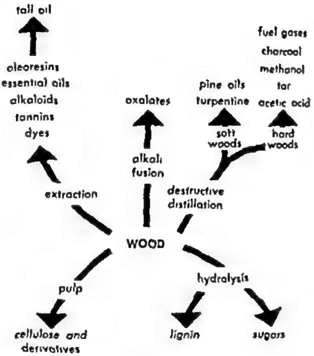
Properties of wood. The structure of wood determines in large part the properties of wood. Thus, a wood with numerous, thick-walled fibers is not only a heavy wood, it is also a hardwood. On the other hand, a wood with large, thin-walled cells is light and soft. A wood with long, interlacing fibers is a strong wood. The durability of wood depends on its chemical composition. See PLANT ANATOMY; WOOD PHYSICS. [O.T.]

Bibliography: See FOREST AND FORESTRY.

Wood chemicals

Wood constitutes the world's most widely used industrial raw material. It ranks third as a fuel and is the source of most of the world's paper and packaging materials and of much of the synthetic fiber.

Chemically, wood is a complex substance, whose cell walls consist essentially of cellulose, other polysaccharides, and lignin. Within the cells and at their surfaces are compounds such as tannins, starches, resins, oils, dyes, alkaloids, and sugars. Wood is a heterogeneous chemical substance, differing markedly from species to species. The accompanying diagram indicates some of the important substances produced from wood.



Important substances produced from wood

Cellulose. About 40–50% of wood is cellulose. Most of the cellulose produced from wood is used in the pulp and paper industry, and much of the remainder is converted to rayon.

The action of acetic anhydride in the presence of sulfuric acid on cellulose ruptures the long-chain molecules and esterifies the fragments, yielding cellobiose octaacetate, which after saponification and hydrolysis with emulsin, yields glucose. Cellobiose is therefore a β-glycoside, and the basic configuration of cellulose is shown in the formula. Measurements of molecular weights of cellulose indicate values of 16,000–2,000,000, the lower values suggesting some degradation in the process of preparation. See CELLOBIOSE; CELLULOSE.

Wood pulp. The sulfate process holds a dominant position in the production of wood pulp or wood cellulose. Logs are debarked, chipped, and digested with aqueous solutions of caustic soda and sodium sulfite. Subsequent washing and bleaching produce a finished cellulose pulp used mainly in paper making. Another widely used method employs calcium bisulfite. A simpler recovery system for the waste liquors is offered by a new method using magnesium bisulfite. See PAPER AND PAPER PRODUCTS.

Rayon. This is a generic term given to synthetic fibers derived from regenerated cellulose. In the viscose process refined pulp is treated with caustic and carbon disulfide to form a water-soluble xanthate.

$$\mathrm{ROH + CS_2 + NaOH \rightarrow [RO\underset{\overset{\|}{S}}{C}-S]^-Na^+ + H_2O}$$

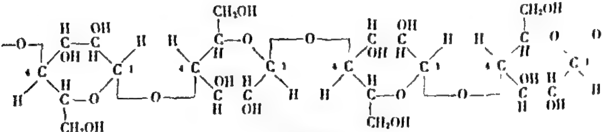
Xanthate

$$\mathrm{[RO\underset{\overset{\|}{S}}{C}-S]^-Na^+ + NaHSO_4 \rightarrow ROH + CS_2 + Na_2SO_4}$$

Viscose process

This viscid liquid is forced through spinnerets into an acid coagulating bath, regenerating the cellulose as filaments. Acetate rayon is made by forcing a solution of cellulose diacetate in acetone through spinnerets into warm air in which the solvent evaporates. See FIBER, MAN-MADE.

Besides rayon, regenerated cellulose is used to prepare derivatives from which are manufactured rayon acetate, plastics, protective coatings, films and foils, cellulose nitrate, explosives, and thickeners.

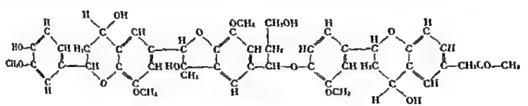


Cellulose

Lignin. This substance is second to cellulose as a principal constituent of wood, averaging about 25%. Its chemical composition in the wood is not clear, but when isolated, lignin seems to have a polymeric structure in which hydroxyl and methoxyl groups are present.

Most of the vast amount of potentially available lignins from wood wastes are used only for fuel. What is used comes principally from the sulfite waste liquors of wood pulp industry. These lignins have been converted to the sulfonic acid derivatives, and are evidently polymers with molecular weights ranging from 2000–15,000.

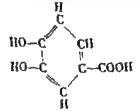
The structure, formation and function of lignin in plants have not yet been defined. When removed from wood, the chemical properties suggest a structure shown below:



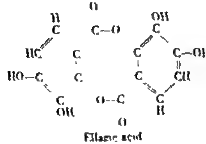
Suggested possible structure for lignin monomer

Lignins find commercial application as binders, fillers, and extenders. Small amounts are converted to vanillin, to tanning materials, and to various dispersants, and are also used to reinforce fillers for rubbers.

Tannins. The tannins constitute an extensive group of water-soluble, complex organic compounds. The principal commercial sources of tannins are the trees: wattle, mangrove, oak, eucalyptus, hemlock, pine, larch, willow, quebracho, chestnut, and urunday. Chips or ground particles of bark or wood are extracted in autoclaves or large open vessels with hot water. As a general rule, the water soluble extract contains about 95% of the tannin available. Chestnut wood for example, contains 5–8% of tannins. Natural tannins exhibit great variations in their chemical constitution and reactions, yet all possess the common property of precipitating gelatin from solution and of forming compounds with collogen and other protein substances in hides to produce leather. The most important use of tannins is in leathermaking. Among the principal identifiable materials isolated from natural tannins are glucose and complexes of glucose with gallic or ellagic acids. *See* TANNIN.



Gallic acid



Ellagic acid

Charcoal. When hardwood is heated in kilns or ovens to final temperatures of 400–500°C, a condensable distillate (pyroligneous acid) is obtained from which acetic acid and methanol are the chief recoverable commercial products. The initial products are nonrecoverable, but utilizable, fuel gases. From one cord of dry hardwood about 140–185 lb of acetic acid and 55–65 lb of *methanol* is recoverable. Of lesser commercial importance are other products including acetone, allyl alcohol, methyl acetate, ethyl acetate, creosote oils, and *pitch*. The residual charcoal, in roughly 30% yields, is not pure carbon. It contains high-boiling tars, complex breakdown products, and some ash. The chief use of charcoal is as a fuel. Some is converted to activated carbon. Today charcoal is economically produced, usually when production takes advantage of the by-products. The yields of acetic acid, charcoal, and methanol depend upon the source of wood and operation efficiency. Resinous soft woods produce distillates of pine tar, *pine oils*, and turpentine. *See* CHARCOAL; PINE OIL; TALL OIL; TURPENTINE.

Alkaloids. The alkaloids are complex nitrogenous compounds, usually basic, usually of vegetable origin, and usually producing some physiological reaction in the animal organism. Many of them are found in the bark, roots, and wood of trees and are of considerable economic importance. *See* ALKALOID.

Oxalic acid. When sawdust is heated with caustic, up to 45% oxalic acid is obtained. The method is obsolete and uneconomical. *See* OXALIC ACID.

Essential oils. Many commercially important essential oils are derived from wood. The following list shows some of the typical oils and their origin. All of them are steam-distilled from the comminuted wood.

Tree	*Essential oil*
Cinnamomum camphora (Sieb.)	Camphor
Juniperus virginiana (L.)	Cedarwood
Santalum album (L.)	Sandalwood
Pinus palustris (Mill.)	Turpentine

See CAMPHOR; ESSENTIAL OILS

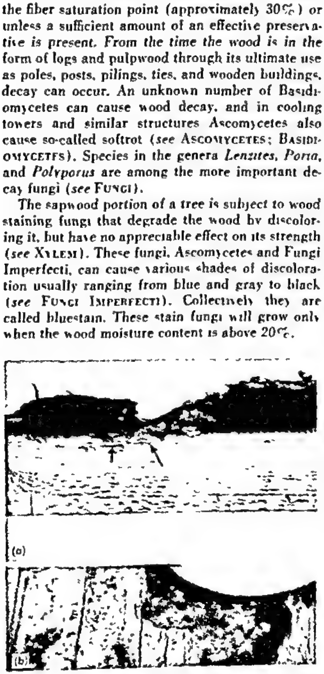
Dyes. Few of the dyes obtained from wood have commercial importance today. They are used principally in the leather industry, and are obtained from the logwood, *Morus tinctoria* (L.), black oak, acacia, and mimosa trees. *See* DYE; LEATHER AND FUR PROCESSING.

Sugars. When wood is treated with a mineral acid, hydrolysis produces a lignin and a liquor containing sugars and other chemicals. The sugars consist of pentoses and hexoses, and can be converted to ethyl alcohol by fermentation. In general, alcohol production by this process is uneconomical. *See* ETHYL ALCOHOL. [E.L.S.]

Wood decay

All woods and wood products are subject to decay if the moisture content of the wood is at or above the fiber saturation point (approximately 30%) or unless a sufficient amount of an effective preservative is present. From the time the wood is in the form of logs and pulpwood through its ultimate use as poles, posts, pilings, ties, and wooden buildings, decay can occur. An unknown number of Basidiomycetes can cause wood decay, and in cooling towers and similar structures Ascomycetes also cause so-called softrot (*see* ASCOMYCETES; BASIDIOMYCETES). Species in the genera *Lenzites, Poria,* and *Polyporus* are among the more important decay fungi (*see* FUNGI).

The sapwood portion of a tree is subject to wood staining fungi that degrade the wood by discoloring it, but have no appreciable effect on its strength (*see* XYLEM). These fungi, Ascomycetes and Fungi Imperfecti, can cause various shades of discoloration usually ranging from blue and gray to black (*see* FUNGI IMPERFECTI). Collectively they are called bluestain. These stain fungi will grow only when the wood moisture content is above 20%.



(a) A section of a utility pole showing extensive decay after only 4 years of service. Arrows indicate the extent of decay. (b) A floor which is badly decayed as a result of the accumulation of moisture in the wood.

Wood deterioration can be prevented by keeping the wood dry. This can be accomplished by using dry wood and proper construction methods. If the wood is to be used in or next to soil or in any other situation where it will periodically have sufficient moisture for fungi to grow, either the wood must be treated with a preservative, or a durable wood must be used. *See* WOOD PRESERVATION.

Some species of wood, such as redwood, cypress, cedars, locust, and chestnut, are naturally resistant to decay and can be used for greenhouse benches and other structures where chemically treated wood might be more expensive or phytotoxic (poisonous to plants). Less durable woods and the sapwood zones of durable species must be protected by preservatives if used where wood cannot be kept dry.

There are a large number of chemicals used as wood preservatives. Oil-soluble preservatives, such as creosote, pentachlorophenol, and copper naphthenate are more suited for products that are going to be in contact with the soil. Water-soluble preservatives have the advantage of being cheaper and less flammable, however, they are more subject to leaching. If a clean paintable surface is desired, pentachlorophenol or some of the water-soluble materials are best.

Wood stains are prevented by dipping the freshly sawn lumber in antistain chemicals containing chlorinated phenols and mercurial compounds. Wood-staining fungi cause an annual loss approaching $10,000,000, while wood losses due to decay are estimated at $300,000,000, not including the cost of preservative and other protective measures. *See* PLANT DISEASE. [D.W.FR.]

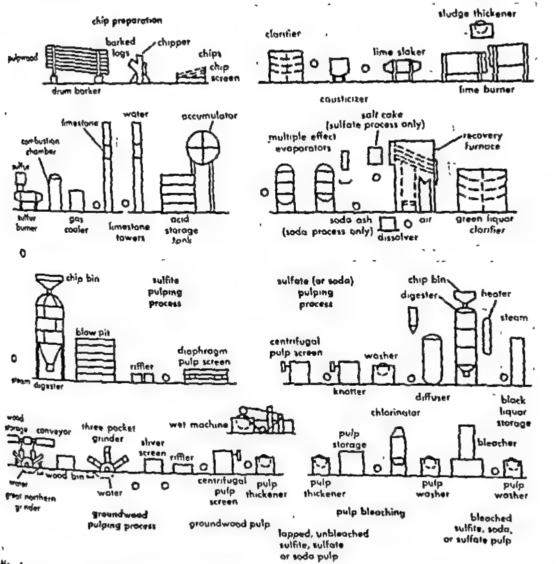
Wood fiber products

Many of the important products of wood are based on the processing of individual wood fibers or groups of fibers. This type of wood use is increasing rapidly and is likely to continue to increase as fiber products are improved and diversified, and as the use of the residues of the forest products industries is extended. *See* WOOD (ANATOMY AND IDENTIFICATION); XYLEM.

Table 1. United States production of wood pulp

Approximate totals (including Alaska)	1957 production, short tons
Wood pulp	21,807,700
Special alpha and dissolving grades	1,008,100
Bleached sulfite	1,981,300
Unbleached sulfite	617,700
Bleached sulfate	3,720,000
Semibleached sulfate	561,000
Unbleached sulfate	7,670,400
Soda	428,100
Groundwood	2,997,300
Semichemical	1,370,800
Defibrated or exploded	1,132,600
Screenings, damaged, and so forth	120,400

SOURCE: U.S. Department of Commerce, Bureau of the Census, *Facts For Industry*, summary for 1957.



Manufacture of wood pulp

Pulping processes. The principal processes for reducing wood to fibers and recombining them are known as pulping processes; these are classified mainly as follows: mechanical (groundwood) and chemical, including the sulfite, sulfate, soda, and semichemical processes.

In the groundwood process, bolts of wood are held against a grindstone and the fibers are ground off in the presence of water (Fig. 1). The yield is 90% or more by weight, but the strength of groundwood pulp is low. To provide adequate strength, 25–50% of pulp produced by the chemical and semichemical processes must be mixed with the groundwood pulp. It is used mostly for newsprint, magazine, catalog, and tissue papers. *See* PAPER AND PAPER PRODUCTS.

The sulfite process employs an acid chemical (calcium, magnesium, sodium, or ammonium bisulfite plus sulfurous acid). The yield is less than half the weight of the wood, but the pulp is much stronger than groundwood pulp. Sulfite pulp is adaptable to a wide variety of uses, including certain grades of book, wrapping, bond, and tissue papers. Purified sulfite pulp is used in viscose rayon and other cellulose derivatives. *See* WOOD CHEMICALS.

The groundwood and sulfite processes are restricted mainly to the pulping of softwoods and only to a limited extent to hardwoods; the sulfate process, however, can be used with almost any wood. Since the chemical used is alkaline (a solution of sodium hydroxide and sodium sulfide), resins, waxes, or fats in the wood do not hinder its pulping action. Hence, it is the process most used for pulping the pines. The yield is less than half the weight of the wood. Sulfate pulp is the strongest of the commercial pulps. Products made from sulfate pulp include high-grade book, magazine, writing, bond, and specialty papers, as well as wrapping, bag paper, and boxboard from unbleached pulps.

The soda process, also alkaline, uses caustic soda

Table 2. Requirements per ton of air-dry pulp

Sulfite		*Soda*	
Wood	1.7–2.2 cords	Wood	1.8–2.6 cords
Sulfur	200–300 lb	New lime	500 lb
Limestone	200–370 lb	Soda ash	250–500 lb
Steam	5000–7500 lb	Steam	4000–6000 lb
Electricity	250 kwhr	Electricity	250 kwhr
Direct labor	4.9 man-hr	Direct labor	5 man-hr
Sulfate		*Groundwood*	
Wood	1.7–2.4	Wood	0.9–1.3 cords
New lime	325 lb	Power	900–1800 kwhr
Salt cake	200–300 lb	Water	35,000–60,000 gal
Steam	2500–5000 lb	Direct labor	2.5–3 man-hr
Electricity	250 kwhr		
Direct labor	5 man-hr		

as the pulping agent; its use is confined principally to the reduction of hardwoods. The yield is about 40–48% of the weight of the wood. Soda pulp is sometimes used alone in making bulky papers, such as blotting paper, where strength is not required. Book, lithograph, and envelope papers are often made from a mixture of sulfite and soda pulps.

The newest of the pulping processes are those classed as a semichemical. They obtain their name from the fact that the wood chips (the fraction with which all but the groundwood process begin) are merely softened by chemical treatment, while the remainder of the pulping action is supplied by a disk attrition mill or by some similar mechanical device for separating the fibers. The chemical solutions may vary. A neutral sodium sulfite solution is most used, although either alkaline sulfate or acid sulfite is applicable. The yield of pulp is of the order of 65–80% of the weight of the wood. The semichemical process is applied predominantly to hardwoods, and the pulps are used in corrugated board for containers, newsprint, magazine papers, and specialty boards.

The most recent of the semichemical processes is known as cold soda pulping because it employs caustic soda without either elevated temperature or pressure. The yields are higher than the other semichemical processes and operating costs are lower. The pulps can be used in the making of corrugated board, newsprint, and book papers.

Wood can be reduced to a coarse fiber product by treating chips with steam or hot water at high temperatures and pressures to weaken fiber bonds, followed by a mechanical fiber separation that results in a mixture of individual fibers and fiber bundles. The yields are 85–95% of the weight of the wood, and the product is used in so-called hardboards, insulating board, roofing felts, and similar products.

Particle board. In addition to the building board produced from coarse fiber, there is an important class of fiber products known as particle board. Particle board consists of wood particles ranging in size from virtually sawdust to large flakes pressed or extruded as sheets and bound by some resin such as urea-formaldehyde or phenol-formaldehyde. Thickness, surface texture, specific gravity, and other properties of particle boards may vary greatly although most boards fall within a weight range of 25–50 lb/ft^3. Faced with veneers or other coverings, particle board is used as co material in furniture and cabinet work and, wit out special surface treatments, as interior pane ing and for exterior paneling in mild exposu conditions.

Bark is processed to a limited extent to produ fillers for plastics, dispersing agents for drillin mud, and other products. See BARK.

Table 3. Production of insulation board, hardboard, and particle board in thousands of square feet*

Year	Insulation *board*	Hard- *board*	Particle *board*	Total
1948	2,386,935	957,080		3,344,0
1956	2,972,593	1,498,193	570,000†	5,040,78
% change	+25	+65		+51

* U.S. Department of Commerce, Bureau of the Censu *Wood Pulp, Paper, and Board Facts for Industry*

† *Estimated*

Constituents and chemical uses. In composi tion, wood is about 50% cellulose, 20% hemicellu loses, and 30% lignin plus a small amount o extractives. Lignin has not been satisfactorily char acterized and its commercial use has not been de veloped in proportion to its abundance. Althoug the total quantity used is small, lignin is the basi of a number of products resulting from the proc essing of residual pulping liquors of the paper in dustry. These products include crude dispersin agents in mud used in oil well-drilling, concrete and ceramics. Vanillin and dimethyl sulfide are als produced, and yeast is cultured on sugars int which a portion of the wood cellulose and hemicel luloses is converted in the pulping process. Tall oil, based on the extractives carried by spent pulping liquors, yields products commonly used in soaps, paints, linoleum, and similar products.

From the standpoint of significant commercial production, however, pulp and paper products constitute the major chemical utilization of wood. The pulp products include special pulps (essentially cellulose) that are dissolved, modified, and regenerated for use in fabrics, plastics, photographic film, nitrocellulose, explosives, cellophane, and similar materials that are produced annually at the rate of well over 900,000 tons. See CELLULOSE; EXPLOSION AND EXPLOSIVE; NITRO AND NITROSO COMPOUNDS; PHOTOGRAPHIC MATERIALS; PLASTICS FABRICATION.

The oldest process for recovering useful chemical products from wood is destructive distillation. At one time this process was used extensively not only for the production of charcoal but also, by processing the resulting crude pyroligneous acid, tar, and wood gases, for the production of methanol (wood alcohol), acetic acid, and related products.

(*see* ACETIC ACID; CHARCOAL; METHANOL). A few plants still operate distillation equipment with by-product recovery, but competition from coal and other raw materials has almost preempted the market (*see* COAL). The major product of destructive distillation is charcoal, produced at the rate of more than a quarter of a million tons in various [illegible] di- [illegible] and rosin from the oleoresin produced by scarifying the longleaf pines of the South. *See* ROSIN; TURPENTINE.

Extracting wood with solvents also has considerable importance. Stumpwood of the longleaf and slash pines is extracted with solvents for the production of rosin and turpentine Extraction of woods and barks with water yields tannins for the leather industry (*see* TANNIN). Most tannin used in the United States is now imported *See* FOREST AND FORESTRY. [D.C.C. (F.P.L.)]

Bibliography: J N. Stephenson (ed), *Pulp and Paper Manufacture Series*, 4 vols, 1950; L. E Wise and E. C. Jahn (eds), *Wood Chemistry*, vols. 1 and 2, 2d ed., 1952

Wood finishing

Any operation performed on wood surfaces aimed at attaining protection against mechanical, physical, and chemical influences or to enhance the natural beauty of the wood Unfinished wood is dimensionally unstable under climatic influences of varying atmospheric moisture and temperature, and is easily marred and soiled (*see* WOOD PHYSICS). Therefore, the main objectives of wood finishing are (1) stabilization—the protective film is not a perfect barrier against moisture, but the rate of moisture exchange into the wood can be retarded; (2) aesthetic quality—the finish should enhance the natural beauty of luster, grain pattern, and color, and (3) sanitation—finishing is intended to close the porous wood surface to reduce contamination by handling, atmospheric grime, and food, and to render it more easily cleaned The chief deteriorating factors against which wood surfaces need to be protected are weathering and dimensional fluctuations due to environmental conditions such as moisture, irradiation, extreme temperature changes, mechanical erosion, or the combined action of these About ¼ in of wood wears away from the exposed surface in a century

Even under indoor conditions, dimensional changes can be considerable due to high or low humidity and to heating or cooling. Wood will swell and shrink as the moisture content and temperature of the surrounding atmosphere change If wood could be covered with a perfectly moisture-impervious coating, dimensional changes could be prevented Since no organic coating is thoroughly impermeable to liquid water or its vapor, only partial protection can be achieved. The more efficient coatings retard moisture movement sufficiently so that under conditions of frequently changing humidity, swelling and shrinking are reduced to tolerable limits.

Stabilizing treatments usually have two aims: to decrease the rate of swelling and shrinking by minimizing the exchange of moisture into or out of the cell wall structure of wood, and to reduce the final equilibrium swelling and shrinking that can take place Surface coatings fall under the first aim.

Surface preparation. The careful preparation of the wood surface is an indispensable requisite for a perfect finishing operation Preparation may include the following steps: machining, rough and fine sanding, watering, and patching These steps might be followed by resanding and, eventually, by bleaching, derosination and destaining, and staining.

Sanding Sanding is usually carried out on a disk, a belt, or on roller sanding machines; usually hand sanding is used only in small-scale operations Two types of coated abrasive paper are used, open and closed coated In the open coated abrasive, 50–70% of the paper surface is covered with abrasive. Closed coated abrasive covers the backing surface completely and is particularly suited to operations involving end-grain wood or wood requiring a higher rate of stock removal and heavy working pressure.

Other factors entering into the selection of the coated abrasive product best adapted for a particular sanding process are: type of wood (hardwood or softwood), amount of stock removal required, speed of belt or drum equipment used, and finish to be obtained. When these factors have been determined, the operation and performance of the selected abrasive is determined by five factors: (1) type, size, and quality of the abrasive grain; (2) backing on the bearer which carries the grain [illegible] able to assembled or shaped pieces; and (5) durability, the quality of the abrasive to maintain sharpness and a fairly constant rate of stock removal for a period of time.

Watering or wash coating. The result of fine sanding can be best evaluated by watering. The surface to be finished is uniformly moistened by spraying, sponging, or short dipping in order to induce fiber raising. The apparently smooth surface swells in the water After removal of excess liquid and drying, the surface will again show unevenness through grain raising In order to impart a desirable brittleness to the raised fibers after drying a solution of animal glue, dextrin, or syn- [illegible]

Patching up This operation is intended to eliminate any damage or undesirable irregularity of the wood surface irrespective of causes. Patching up is accomplished either by partial removal of the damage such as a knot or scar, and replacement by a

Table 2. Requirements per ton of air-dry pulp

Sulfite		*Soda*	
Wood	1.7–2.2 cords	Wood	1 8–2 6 cords
Sulfur	200–300 lb	New lime	500 lb
Limestone	200–370 lb	Soda ash	250–500 lb
Steam	5000–7500 lb	Steam	4000–6000 lb
Electricity	250 kwhr	Electricity	250 kwhr
Direct labor	4 9 man-hr	Direct labor	5 man-hr
Sulfate		*Groundwood*	
Wood	1 7–2 4	Wood	0 9–1 3 cords
New lime	325 lb	Power	900–1800 kwhr
Salt cake	200–300 lb	Water	35,000–60,000 gal
Steam	2500–5000 lb	Direct labor	2.5–3 man-hr
Electricity	250 kwhr		
Direct labor	5 man-hr		

as the pulping agent; its use is confined principally to the reduction of hardwoods. The yield is about 40–48% of the weight of the wood Soda pulp is sometimes used alone in making bulky papers, such as blotting paper, where strength is not required. Book, lithograph, and envelope papers are often made from a mixture of sulfite and soda pulps.

The newest of the pulping processes are those classed as a semichemical. They obtain their name from the fact that the wood chips (the fraction with which all but the groundwood process begin) are merely softened by chemical treatment, while the remainder of the pulping action is supplied by a disk attrition mill or by some similar mechanical device for separating the fibers. The chemical solutions may vary. A neutral sodium sulfite solution is most used, although either alkaline sulfate or acid sulfite is applicable. The yield of pulp is of the order of 65–80% of the weight of the wood. The semichemical process is applied predominantly to hardwoods, and the pulps are used in corrugated board for containers, newsprint, magazine papers, and specialty boards

The most recent of the semichemical processes is known as cold soda pulping because it employs caustic soda without either elevated temperature or pressure. The yields are higher than the other semichemical processes and operating costs are lower. The pulps can be used in the making of corrugated board, newsprint, and book papers.

Wood can be reduced to a coarse fiber product by treating chips with steam or hot water at high temperatures and pressures to weaken fiber bonds, followed by a mechanical fiber separation that results in a mixture of individual fibers and fiber bundles. The yields are 85–95% of the weight of the wood, and the product is used in so-called hardboards, insulating board, roofing felts, and similar products.

Particle board. In addition to the building board produced from coarse fiber, there is an important class of fiber products known as particle board. Particle board consists of wood particles ranging in size from virtually sawdust to large flakes pressed or extruded as sheets and bound by some resin such as urea-formaldehyde or phenol-formaldehyde. Thickness, surface texture, specific gravity, and other properties of particle boards may vary greatly although most boards fall within a weight range of 25–50 lb/ft^3. Faced with veneers or other coverings, particle board is used as core material in furniture and cabinet work and, without special surface treatments, as interior paneling and for exterior paneling in mild exposure conditions.

Bark is processed to a limited extent to produce fillers for plastics, dispersing agents for drilling mud, and other products. *See* BARK.

Table 3. Production of insulation board, hardboard, and particle board in thousands of square feet*

Year	Insulation board	Hard-board	Particle board	Total
1948	2,386,955	957,080		3,344,035
1956	2,972,593	1,498,193	570,000†	5,040,786
% change	+25	+65		+51

* U S Department of Commerce, Bureau of the Census, *Wood Pulp, Paper, and Board Facts for Industry*.
† Estimated

Constituents and chemical uses. In composition, wood is about 50% cellulose, 20% hemicelluloses, and 30% lignin plus a small amount of extractives Lignin has not been satisfactorily characterized and its commercial use has not been developed in proportion to its abundance. Although the total quantity used is small, lignin is the basis of a number of products resulting from the processing of residual pulping liquors of the paper industry These products include crude dispersing agents in mud used in oil well-drilling, concrete, and ceramics Vanillin and dimethyl sulfide are also produced, and yeast is cultured on sugars into which a portion of the wood cellulose and hemicelluloses is converted in the pulping process Tall oil, based on the extractives carried by spent-pulping liquors, yields products commonly used in soaps, paints, linoleum, and similar products.

From the standpoint of significant commercial production, however, pulp and paper products constitute the major chemical utilization of wood. The pulp products include special pulps (essentially cellulose) that are dissolved, modified, and regenerated for use in fabrics, plastics, photographic film, nitrocellulose, explosives, cellophane, and similar materials that are produced annually at the rate of well over 900,000 tons. *See* CELLULOSE; EXPLOSION AND EXPLOSIVE; NITRO AND NITROSO COMPOUNDS; PHOTOGRAPHIC MATERIALS; PLASTICS FABRICATION.

The oldest process for recovering useful chemical products from wood is destructive distillation. At one time this process was used extensively not only for the production of charcoal but also, by processing the resulting crude pyroligneous acid, tar, and wood gases, for the production of methanol (wood alcohol), acetic acid, and related products

... A few
... d
other raw materials has almost preempted the market (see COAL). The major product of destructive distillation is charcoal, produced at the rate of more than a quarter of a million tons in various types of nonrecovery kilns for uses about evenly divided between smelting and cookery.

Other important products are turpentine and rosin from the oleoresin produced by scarifying the longleaf pines of the South. See ROSIN; TURPENTINE.

Extracting wood with solvents also has considerable importance. Stumpwood of the longleaf and slash pines is extracted with solvents for the production of rosin and turpentine. Extraction of woods and barks with water yields tannins for the leather industry (see TANNIN). Most tannin used in the United States is now imported. See FOREST AND FORESTRY. [D.G.C. (F.P.L.)]

Bibliography: J. N. Stephenson (ed.), *Pulp and Paper Manufacture Series*, 4 vols., 1950; L. E. Wise and E. C. Jahn (eds.), *Wood Chemistry*, vols. 1 and 2, 2d ed., 1952.

Wood finishing

Any operation performed on wood surfaces aimed at attaining protection against mechanical, physical, and chemical influences or to enhance the natural beauty of the wood. Unfinished wood is dimensionally unstable under climatic influences of varying atmospheric moisture and temperature, and is easily marred and soiled (see WOOD PHYSICS). Therefore, the main objectives of wood finishing are (1) stabilization—the protective film is not a perfect barrier against moisture, but the rate of moisture exchange into the wood can be retarded, (2) aesthetic quality—the finish should enhance the natural beauty of luster, grain pattern, and color, and (3) sanitation—finishing is intended to close the porous wood surface to reduce contamination by handling, atmospheric grime and food, and to render it more easily cleaned. The chief deteriorating factors against which wood surfaces need to be protected are weathering and dimensional fluctuations due to environmental conditions such as moisture, irradiation, extreme temperature changes, mechanical erosion, or the combined action of these. About ¼ in. of wood wears away from the exposed surface in a century.

Even under indoor conditions, dimensional changes can be considerable due to high or low humidity and to heating or cooling. Wood will swell and shrink as the moisture content and temperature of the surrounding atmosphere change. If wood could be covered with a perfectly moisture-impervious coating, dimensional changes could be prevented. Since no organic coating is thoroughly impermeable to liquid water or its vapor, only partial protection can be achieved. The more efficient coatings retard moisture movement sufficiently so that under conditions of frequently changing humidity, swelling and shrinking are reduced to tolerable limits.

Stabilizing treatments usually have two aims: to decrease the rate of swelling and shrinking by minimizing the exchange of moisture into or out of the cell wall structure of wood, and to reduce the final equilibrium swelling and shrinking that can take place. Surface coatings fall under the first aim.

Surface preparation. The careful preparation of the wood surface is an indispensable requisite for a perfect finishing operation. Preparation may include the following steps: machining, rough and fine sanding, watering, and patching. These steps might be followed by resanding and, eventually, by bleaching, derosination and destaining, and staining.

Sanding. Sanding is usually carried out on a disk, a belt, or on roller sanding machines; usually hand sanding is used only in small-scale operations. Two types of coated abrasive paper are used, open and closed coated. In the open coated abrasive, 50–70% of the paper surface is covered with abrasive. Closed coated abrasive covers the backing surface completely and is particularly suited to operations involving end-grain wood or wood requiring a higher rate of stock removal and heavy working pressure.

Other factors entering into the selection of the coated abrasive product best adapted for a particular sanding process are: type of wood (hardwood or softwood), amount of stock removal required, speed of belt or drum equipment used, and finish to be obtained. When these factors have been determined, the operation and performance of the selected abrasive is determined by five factors: (1) type, size, and quality of the abrasive grain; (2) backing on the bearer which carries the grain into action, (3) bonding, the process by which the abrasive is bonded to the backing, (4) flexibility, the characteristic which makes the product adaptable to assembled or shaped pieces, and (5) durability, the quality of the abrasive to maintain sharpness and a fairly constant rate of stock removal for a period of time.

Watering or wash coating. The result of fine sanding can be best evaluated by watering. The surface to be finished is uniformly moistened by spraying, sponging, or short dipping in order to induce fiber raising. The apparently smooth surface swells in the water. After removal of excess liquid and drying, the surface will again show unevenness through grain raising. In order to impart a desirable brittleness to the raised fibers after drying, a solution of animal glue, dextrin, or synthetic water-soluble resin of the urea-formaldehyde type is used. Removal of the raised grain by sanding must be done with extreme care, especially on thin veneer.

Patching up. This operation is intended to eliminate any damage or undesirable irregularity of the wood surface irrespective of causes. Patching up is accomplished either by partial removal of the damage, such as a knot or scar, and replacement by a

sound wood-patch or dowel, by gluing, or simply by filling in with a puttylike material such as plastic wood.

Derosination and destaining. Certain species contain excessive amounts of rosin which may interfere with the drying of varnishes, repel the stain used, or cause a patchy penetration of the stain Derosination is effected through saponification with alkaline aqueous solutions or organic solvents. A combination of both in the form of acetone with added ammonia or sodium carbonate solution is efficient in difficult cases. Destaining is sometimes required to remove spotty stains caused by fungal contamination which occurred during drying or seasoning (*see* WOOD DECAY; WOOD PRESERVATION). Destaining is similar or identical to methods for bleaching on limited areas.

Bleaching. Chemically, bleaching consists of a decomposition of wood colors. Frequently wood is more successfully finished in lighter and brighter shades than in its natural coloring. In such instances, darker-colored woods are lightened by bleaching them before applying stain without obscuring their attractive texture. The most important chemicals used in bleaches are hydrogen peroxide, sodium bisulfite, sodium hyposulfite, sodium perborate, oxalic acid, potassium permanganate, and sodium or calcium hypochlorite. The active ingredients are applied by brushing, spraying, sponging, or dipping, usually in combination with some secondary chemical such as ammonia or sodium hydroxide Residual chemicals have to be carefully removed from the wood. Commercial bleaching compounds are available. *See* BLEACHING

Staining Staining may be defined as the application of coloring matter to wood surfaces. The materials commonly used as stains consist of transparent or semitransparent colors suspended in a solution of volatile liquids which usually incorporate a small amount of a nonvolatile binder to facilitate spreading, penetration, and fixation of color Stains are intended more to impart color effects without obscuring grain pattern than to cover with a protective coating. Staining serves three vital functions in wood finishing: (1) it accentuates the color contrast of the grain pattern, (2) it evens up or accentuates color differences, producing a more uniform over-all tone; and (3) it imitates woods of greater appeal. The following types of stains are available (1) water-soluble, (2) oil, (3) pigmented, (4) alcohol-soluble, and (5) nongrain-raising stains; all are used with various organic solvents. *See* SURFACE COATING.

Grain-raising stains have to be refinished by light sanding. Aniline dyes are precipitated on and within the fibers and as a rule are not water-resistant Alkaline organic dyes are resistant to light but nonresistant to acids.

Prefinishing. Priming, filling, and sealing are the major steps in prefinishing.

Priming (exterior). Two types can be distinguished: primers which form a coat or film on the surface; and penetrant primers which usually combine water repellents, synthetic resins, and fung cidal preparations in an organic solvent. Penetrar primers promote deeper penetration than do coa type primers.

Filling (interior). In high-grade interior finis ing, fillers are used to produce a smooth, unifor surface for subsequent varnishing or lacquerin (*see* LACQUER; VARNISH). Open-grained woods re quire filling; others are usually finished withou filling. Fillers can be combined with stains or pig mented to emphasize the grain pattern. Chemically fillers are colorless or covering pigments that ar dispersed in a vehicle consisting of drying oils alkyd or other synthetic resins, and proper thinner based on organic solvents. Fillers are available i liquid or paste form for close- and open-graine woods and are applied by brushing, spraying, o wiping.

Sealing (interior) or wash coating. Sealin serves the following purposes: (1) it prevents pen etration of or redissolution in subsequent finis coats; (2) it seals in stain and filler; and (3) i imparts brittleness to raised grain The sealin coat can be applied after staining and before o after filling.

Materials and processes. Finishing composition consist of the vehicle, which is also called the con tinuous or external phase, and the pigment, calle the discontinuous or internal phase. The vehicl consists of natural and synthetic resins, waxes, and drying oils, and aqueous or organic solvents. The internal phase contains inorganic pigments, or ganic dyestuffs, lakes (organic dyes precipitated on inorganic bearers); it might also contain fillers and extenders. Internal or external plasticization is used to reduce brittleness of the external phase. The film formative process is either physical, effected by evaporation as in organic solvents, hydrosols, emulsions, and lattices; or chemical, by air drying as in oil vehicles, oleo-resinous varnishes, oxygen-convertible polymers, or catalyzed synthetics, or by drying at elevated temperatures. The most commonly used pigments are white lead, zinc oxide, and titanium dioxide, together with colors for tinting. *See* PAINT.

Kinds of wood finishes. Wood finishes can be grouped in two main classes: exterior and interior. Exterior finishes are expected to resist chiefly climatic influences, sometimes combined with mechanical abuse, as in outdoor furniture. Interior finishes are expected to fill such exacting standards of appearance and quality as sheen, luster, gloss, or chemical and abrasion resistance. Water resistance is also required in exposed interior locations, such as kitchens, bathrooms, and verandas.

Exterior finishes include: (1) transparent natural finishes showing wood texture, which are further classified as (*a*) surface coatings (transparent finishes consisting of varnishes or synthetics), and (*b*) penetrating finishes including stains, oils, sealers, and water repellents or fungicides; (2) pigmented covering paints, applied on a prim ing coat containing white lead but no zinc oxide

and having a total pigment the same as the finish-coat. These covering paints must contain sufficient thinner (solvents) to wet wood quickly and adequate binders (bodied oils and resins) to prevent undue penetration of vehicle into the wood.

Interior finishes include: (1) transparent penetrating finishes consisting of stains in combination with water repellent finishes based on waxes and other nonvolatile ingredients; (2) transparent film-forming coating based on shellac, nitrocellulose lacquers, or oleoresinous varnishes in combination with stains, fillers, and sealers; and (3) pigmented covering paints (enamels) used in combination with a priming coat.

Rubbing and polishing. Rubbing and polishing are the final steps and are intended to produce certain surface effects on the coating film. Rubbing is the application of an abrasive in conjunction with a lubricant to level, dull, or deaden luster; polishing is similar and denotes adjustment of sheen or luster, sometimes applying very fine soft abrasives or only waxy ingredients if the purpose is to increase gloss. Lubricants used are water and soap or detergents, petroleum based thin oils, and oil-in-water type emulsions. Abrasives used are pumice, tripoli, diatomaceous earth, and rottenstone (*see* DIATOMACEOUS EARTH; PUMICE). Natural and synthetic wax compositions including abrasives are available for polishing in solid, paste, or liquid forms.

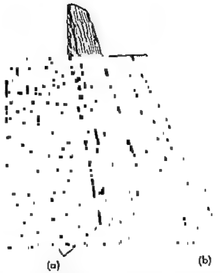
Methods of application. The most commonly used methods are hand brushing, spraying, and hot spraying. Newer methods for flat surfaces are roller and flow coating for walls, furniture, and door panels, and dipping for small parts such as toys. Drying or curing is accelerated by hot air or radiant heat produced by infrared lamps.

Finishing properties of wood. The suitability of a wood for a certain purpose is determined by density, mean pore size, springwood summerwood distribution, grain orientation (flat or plain sawn, or quarter sawn), moisture content, defects such as knots and pitch pockets, and extractives such as resin, oils, tannins, and polyphenols. In general, the greater its density, the harder the wood. As a class, the softwoods are more abundant, cheaper, more easily worked, and better suited for the exteriors of buildings than are the hardwoods. The best woods for interior painting are hardwoods. Hardwoods with pores larger than those of birch are not well adapted to finishing with exterior surface coatings because the coating does not level and fill the large pores. *See* BIRCH; WOOD (ANATOMY AND IDENTIFICATION); XYLEM.

The commercially important native woods have been classified on the basis of finish-holding characteristics into five groups:

Group 1. Cedars, baldcypress, and redwood, all softwoods. Woods of this group satisfactorily hold all of the common kinds of house and barn paints.

Group 2. Eastern white pine, sugar pine, and western white pine, all softwoods. Paints containing zinc oxide, if applied directly to the woods of



Two examples of boards cut from a log, showing surface and grain patterns. (a) Quarter-sawed (radially sawed) board. (b) Plain-sawed (tangentially sawed) board. (From H. J. Fuller and O. Tippo, College Botany, rev. ed., Holt, 1954)

this group, are slightly less durable, but other paints are as durable as on woods of group 1.

Group 3. White fir, hemlocks, ponderosa pine, and spruces of the softwoods, aspen, basswood, cottonwood, magnolia, and yellow poplar of the hardwoods. Paints deteriorate somewhat more rapidly on woods of this group than on woods of groups 1 and 2. On ponderosa pine, paints containing zinc oxide are less durable than paints free from it.

Group 4. Douglas fir, southern yellow pine, and western larch of the softwoods, beech, birch, gums, and maples of the hardwoods. Paints deteriorate more rapidly on woods of this group than on woods of groups 1, 2, and 3. On southern yellow pine, paints containing zinc oxide are less durable than paints free from it.

Group 5. Ash, chestnut, elm, hickory, oak, and walnut, all hardwoods with large pores. Woods of this group require wood filler to fill the pores before smooth coatings of paint or enamel can be applied. See separate articles listed by common names for all trees mentioned above.

Interior and exterior finishing. Each of these types of finishing involves specific steps.

Interior finishing. A complete representative schedule consists of the following steps starting from the conclusion of the sanding operation: (1) watering, (2) bleaching, (3) staining, (4) filling, (5) sealing, (6) spraying base coat, (7) spraying finish coat, (8) rubbing, (9) cleaning, and (10) polishing. The number and sequence of these basic steps may be varied considerably, however, to meet specific requirements of proprietary (trade) finishing materials, processes, technical quality, and aesthetic appearance. The trend in interior finishing of wood is to emphasize the sensorial and tactile qualities of wood.

sound wood-patch or dowel, by gluing, or simply by filling in with a puttylike material such as plastic wood.

Derosination and destaining. Certain species contain excessive amounts of rosin which may interfere with the drying of varnishes, repel the stain used, or cause a patchy penetration of the stain. Derosination is effected through saponification with alkaline aqueous solutions or organic solvents. A combination of both in the form of acetone with added ammonia or sodium carbonate solution is efficient in difficult cases. Destaining is sometimes required to remove spotty stains caused by fungal contamination which occurred during drying or seasoning (*see* WOOD DECAY; WOOD PRESERVATION). Destaining is similar or identical to methods for bleaching on limited areas.

Bleaching. Chemically, bleaching consists of a decomposition of wood colors. Frequently wood is more successfully finished in lighter and brighter shades than in its natural coloring. In such instances, darker-colored woods are lightened by bleaching them before applying stain without obscuring their attractive texture. The most important chemicals used in bleaches are hydrogen peroxide, sodium bisulfite, sodium hyposulfite, sodium perborate, oxalic acid, potassium permanganate, and sodium or calcium hypochlorite. The active ingredients are applied by brushing, spraying, sponging, or dipping, usually in combination with some secondary chemical such as ammonia or sodium hydroxide. Residual chemicals have to be carefully removed from the wood. Commercial bleaching compounds are available. *See* BLEACHING.

Staining. Staining may be defined as the application of coloring matter to wood surfaces. The materials commonly used as stains consist of transparent or semitransparent colors suspended in a solution of volatile liquids which usually incorporate a small amount of a nonvolatile binder to facilitate spreading, penetration, and fixation of color. Stains are intended more to impart color effects without obscuring grain pattern than to cover with a protective coating. Staining serves three vital functions in wood finishing: (1) it accentuates the color contrast of the grain pattern, (2) it evens up or accentuates color differences, producing a more uniform over-all tone, and (3) it imitates woods of greater appeal. The following types of stains are available: (1) water-soluble, (2) oil, (3) pigmented, (4) alcohol-soluble, and (5) nongrain-raising stains, all are used with various organic solvents. *See* SURFACE COATING.

Grain-raising stains have to be refinished by light sanding. Aniline dyes are precipitated on and within the fibers and as a rule are not water-resistant. Alkaline organic dyes are resistant to light but nonresistant to acids.

Prefinishing. Priming, filling, and sealing are the major steps in prefinishing.

Priming (exterior). Two types can be distinguished: primers which form a coat or film on the surface, and penetrant primers which usually combine water repellents, synthetic resins, and fung cidal preparations in an organic solvent. Penetrai primers promote deeper penetration than do coa type primers.

Filling (interior). In high-grade interior finisl ing, fillers are used to produce a smooth, uniforı surface for subsequent varnishing or lacquerin (*see* LACQUER; VARNISH). Open-grained woods re quire filling; others are usually finished withou filling. Fillers can be combined with stains or pig mented to emphasize the grain pattern. Chemicall fillers are colorless or covering pigments that aı dispersed in a vehicle consisting of drying oil alkyd or other synthetic resins, and proper thinner based on organic solvents. Fillers are available i liquid or paste form for close- and open-graine woods and are applied by brushing, spraying, o wiping.

Sealing (interior) or wash coating. Sealin serves the following purposes: (1) it prevents per etration of or redissolution in subsequent finis coats, (2) it seals in stain and filler; and (3) i imparts brittleness to raised grain. The sealin coat can be applied after staining and before o after filling.

Materials and processes. Finishing composition consist of the vehicle, which is also called the con tinuous or external phase, and the pigment, callec the discontinuous or internal phase. The vehicl consists of natural and synthetic resins, waxes, anc drying oils, and aqueous or organic solvents. Th internal phase contains inorganic pigments, or ganic dyestuffs, lakes (organic dyes precipitatec on inorganic bearers); it might also contain filler and extenders. Internal or external plasticization i used to reduce brittleness of the external phase The film formative process is either physical, ef fected by evaporation as in organic solvents, hy drosols emulsions, and lattices; or chemical, by air drying as in oil vehicles, oleo resinous var nishes, oxygen-convertible polymers, or catalyzed synthetics, or by drying at elevated temperatures. The most commonly used pigments are white lead, zinc oxide, and titanium dioxide, together with col ors for tinting. *See* PAINT.

Kinds of wood finishes. Wood finishes can be grouped in two main classes: exterior and interior. Exterior finishes are expected to resist chiefly climatic influences, sometimes combined with mechanical abuse, as in outdoor furniture. Interior finishes are expected to fill such exacting standards of appearance and quality as sheen, luster, gloss, or chemical and abrasion resistance. Water resistance is also required in exposed interior locations, such as kitchens, bathrooms, and verandas.

Exterior finishes include: (1) transparent natural finishes showing wood texture, which are further classified as (*a*) surface coatings (transparent finishes consisting of varnishes or synthetics), and (*b*) penetrating finishes including stains, oils, sealers, and water repellents or fungicides; (2) pigmented covering paints, applied on a prim ing coat containing white lead but no zinc oxide

...scopic water attains moisture equilibrium with the environment. This equilibrium moisture content is a function of relative vapor pressure and temperature (Fig. 1). It also varies with type of wood, extractive content, and exposure history.

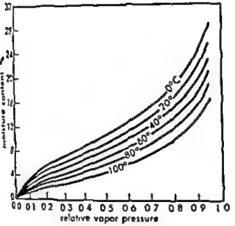


Fig. 1. Representative equilibrium moisture content curves for initial desorption of wood at various temperatures. (From H. D. Tiemann, Wood Technology, 3d ed., Pitman, 1951)

Heat of wetting. The additional heat required, above the heat of vaporization of free water, to evaporate water from wood is called heat of wetting. It varies from about 280 calories per gram of water evaporated for oven-dry wood to essentially zero at the fiber saturation point and above.

Shrinking and swelling. The dimensional change associated with loss or gain of bound water is defined respectively as shrinking or swelling. Cell wall shrinkage is somewhat less than the volume of water desorbed.

Volumetric shrinkage. Owing to the uncertain behavior of the cell cavities, volumetric shrinkage of the gross wood cannot be predicted accurately. The cell cavities may shrink, remain constant, or swell when the cell walls lose moisture, depending on the particular wood and on the drying technique. On the average, the cavities remain essentially constant in size, as a result of the restraining effects of the inner and outer layers of the secondary cell wall. The per cent volumetric shrinkage S_f from green to oven-dry for a given wood can be estimated by use of the equation $S_f = 0.9\ G_f\ M_f$, where G_f is the nominal green specific gravity based on oven-dry weight and green volume, and M_f is the fiber saturation point.

Differential shrinkage. Longitudinal shrinkage in normal mature wood is almost negligible, since the crystallites are oriented almost longitudinally and shrinkage occurs perpendicular to these. Tangential shrinkage is about twice radial shrinkage, owing to the action of at least two mechanisms. One is the restraining effect of the rays on radial shrinkage. The other is the high volumetric shrinkage and strength of latewood compared with earlywood. See ELASTICITY.

Swelling pressure. The approximate swelling pressure p, exerted when the cell wall is restrained in the swelling directions and the surrounding vapor pressure is increased from h_1 to h_2, is given by the Katz equation, $p = (\rho RT/m)\ \ln\ (h_2/h_1)$, where R, T, ρ, and m are the gas constant, Kelvin temperature, sorbed water density, and molecular weight, respectively. The pressure p cannot be realized in the gross wood unless the cell cavities, into which the cell walls normally are forced to swell if externally restrained, are completely collapsed. See OSMOSIS; PLANT, WATER RELATIONS OF.

Moisture movement. Water moves through wood by three distinct mechanisms: capillary movement through the cell cavities; hygroscopic or bound-water movement through the cells walls; and vapor movement through the cell cavities.

Capillary movement. Water movement of this type occurs above the fiber saturation point, and is motivated by capillary tensions in the cell-cavity water. Evaporation of moisture from excised cell cavities near the surface of a drying sample causes air-water menisci to retreat toward the tapering cell-cavity ends or into pit membrane openings. The increased capillary tensions, thus induced by reduction in meniscus radius, pull on the water contained in the interior cell cavities. If this water is free to move, that is, if air bubbles are present to expand as water is removed, it migrates to the surface and evaporates. In some cases, individual cells or cell groups are completely water-filled and have access to air only through tiny apertures with menisci small enough to support large capillary forces. The high tensions thus exerted on the cell walls may be sufficient to cause their collapse, particularly at the higher temperatures where the wood is weakened. See LUMBER MANUFACTURE.

Bound-water movement. Water movement of this type through the cell walls results from activated diffusion at low moisture contents. Individual water molecules migrate from one sorption site (hydroxyl group) to another in the direction of increasing numbers of unoccupied sites, that is, toward regions of lower moisture content. At higher moisture contents, bound water diffuses from high to low concentration regions more nearly like diffusion in solutions of unequal concentration.

Vapor movement. The driving force for vapor movement through the cell cavities is the gradient of vapor pressure, and the movement obeys the laws for vapor diffusion through small capillaries (see VAPOR PRESSURE). Its relative importance increases with temperature owing to the strong increase of water-vapor pressure with temperature.

Thermal properties of wood. All physical properties are temperature dependent, and these effects are discussed under individual headings. Only heat transmission and expansion are considered here.

Thermal conductivity. The thermal conductivity, K, across the grain is only about 40% of the parallel-to-grain value. It is sometimes a few per cent

Exterior finishing. A complete scheme for this kind of finishing consists of (1) fungicide treatment, (2) water-repellent treatment, (3) patching up defects, (4) primer coat, (5) stain coat, and (6) finish coat. However, the number and sequence of steps change for specific purposes. For example, the stain coat for pigmented paints or finish coat for transparent finishes may be omitted; likewise primer, water repellent, and fungicides can be combined.

Special methods. In addition to the conventional practices described, the following methods are employed in special cases: (1) improvement of the surface texture for painting of low-grade plywood or hardboard lumber, achieved by the application of resin impregnated Kraft paper; (2) decorative and protective finishes, produced by a similar method replacing conventional methods of finishing; (3) factory semiprefinishing of composite panel materials, for example, core boards, plywood, reconstituted wood panels, and parquet flooring; (4) production of metal-clad wood as a constructional material; (5) use of aluminum foil with colored organic enamels on wood parts, replacing finishing and insuring perfect moisture impermeability; (6) flame spraying of metals and plastics; (7) use of high vacuum metalizing on wood, mostly on small parts; and (8) burnish finishing, an application of frictional heat and pressure with cork-covered continuous belts on a sander-type machine.

Finishing preservative-treated wood. Wood treated with water-soluble preservatives can be finished satisfactorily after proper redrying. Coal-tar creosote, or other dark oily preservatives such as chlorinated tars, can be painted if weathered for a considerable time before painting. However, discoloration may occur. Wood preserved with solvent-borne chlorinated phenols can be painted without difficulty, provided the solvents are thoroughly evaporated from the surfaces.

Evaluation and testing. The weather-resisting qualities (paint holding) of exterior finishes are most reliably evaluated through long-term exposure of panels on test fences in different climates. Laboratory artificial weathering tests simulating rain (spraying), sunshine (ultraviolet radiation), and seasoning (drying at elevated temperatures) under recycling are in use as a more rapid method of evaluation. Through comparison of results with a standard of known weather-resisting quality, fairly reliable relationships can be established in each instance between the effect of natural and artificial weathering. Mildew resistance is tested in bioassays, and color permanence is determined by visual or colorimetric comparison of standards.

Mar-resistance, permanence of color, continuity of coating, and adhesion (durability), together with abrasive resistance and chemical resistance in flooring or furniture, such as table tops, are the most important consumer considerations.

Working characteristics in spraying, drying, leveling, rate of cure (tackiness), hardness, covering power, and viscosity (flowing characteristics) are tested with the aid of special laboratory facilities. Recently the interactions between finishing compositions and wood structures are being studied by means of the microscope and other physical and chemical methods. The significance of surface texture is considered in its relationship to wood finishing operations. *See* FOREST AND FORESTRY.

[J.E.M.]

Bibliography: Am. Soc. Testing Materials, *ASTM Standards*, 1955; A. V. Blom, *Organic Coatings in Theory and Practice*, 1949; *List of Publications on Wood Finishing Subjects*, Forest Prod. Lab No 454, 1956; R. Scharff, *Complete Book of Wood Finishing*, 1956; *Wood Handbook*, USDA No. 72, 1955.

Wood physics

That area of wood science concerned with the physical and mechanical properties of wood, and the factors which affect them. Since wood consists of aggregates of long tubular cells most of which are oriented longitudinally in the tree, whereas the ray cells are oriented radially, wood is an orthotropic material and exhibits different physical behavior in the three main structural directions, longitudinal, radial, and tangential. The orientation of most of the cellulose crystallites is also longitudinal. *See* CELLULOSE; WOOD (ANATOMY AND IDENTIFICATION); XYLEM.

Wood density. The cell wall density for most woods is essentially the same, that is, about 1.5 g/cm^3 in the oven-dry condition and about 1.4 g cm^3 at the fiber-saturation point (*see* CELL WALLS IN PLANTS). The main factor governing the density of a given wood sample is the ratio of cell-wall to cell-cavity volume. Variations in extraneous compounds in the cell walls and cavities also affect density, as does the moisture content. In many trees the wood formed early in the growing season, earlywood or springwood, is less dense than wood formed later, latewood or summerwood. The term nominal specific gravity G when used for wood in the United States is defined as the ratio of the oven-dry weight of a wood sample to the weight of water displaced by its volume at some specified moisture content.

Hygroscopicity. Wood under normal atmospheric conditions contains water which can be removed by drying. The total moisture content M is usually defined as the ratio of the weight of water lost, by drying in an oven at 100–105°C, to the oven-dry weight of the wood, expressed as a percentage. Green wood normally contains water in the cell cavities as well as in the cell walls. The cavity, or free, water can be removed by exposure to relative humidities only slightly lower than 100%, or about 99.9%. The cell-wall, or bound, water is hygroscopic and remains in the wood unless exposed to humidities lower than 100%. The cell-wall moisture content M_f, usually 25–30%, at which the cell cavities are empty but the walls are saturated, is defined as the fiber saturation point. The hygro-

in wood density increases the cell-wall thickness which greatly increases resistance to these bending moments, approximately as the square of the wall thickness or wood density. Cell wall stresses induced by longitudinal forces are largely extensional in nature, and the resistance to these increases almost linearly with wall thickness or density

Wood strength at temperatures above freezing is not affected by moisture content changes above the fiber-saturation point. In the hygroscopic range however, most strength properties increase exponentially with decreasing moisture content, as indicated by the relationship, $S_m = S_o e^{-bM}$, where S_m and S_o are the strength values at the moisture content M and oven-dry, and b is an empirical constant dependent upon the particular strength property (usually $b = 0.01$ to 0.03). The shock resistance or toughness of wood remains almost constant or may actually decrease with decreasing moisture content owing to the greater pliability of wood at the higher moisture contents

Temperature affects wood strength in two ways: the immediate or reversible effect and the permanent or irreversible effect due to thermal degradation.

The immediate effect of temperature causes a decrease in strength with increasing temperature. This effect is more pronounced as moisture content M increases. For example, the per cent change B, in the modulus of rupture in bending, per degree C from its value at 20°C, over the temperature range from −20°C to +70°C, is given approximately by

$$B = 0.22 + 0.34M$$

in the hygroscopic range

The permanent effect of temperature is caused by thermal degradation of wood exposed to high temperatures, resulting in weight and strength loss when the wood is returned to standard conditions A linear relation between the logarithm of strength or weight loss and heating time is found when wood is maintained at a constant elevated temperature The slope of this linear relationship increases with temperature and when plotted against the reciprocal of absolute temperature, another linear curve results, from which the activation energy for the degradative process can be calculated A J. Stamm found energies of 16 and 28 kcal/mole for wood heated in steam and in dry air, respectively. The practical results of strength tests, as related to heating time and temperature, indicate that negligible loss in strength results from long periods of wood exposure to ordinary room temperatures or from short exposure periods to the high temperatures used in kiln drying

The effects of defects on wood strength are difficult to evaluate quantitatively, but their effects are known qualitatively Knots reduce tensile strength but may have little effect on the compressive strength Checks or cracks considerably reduce the parallel to-grain shear strength Decay, caused by wood-destroying fungi, reduces all strength properties, particularly shock resistance (*see* WOOD DECAY; WOOD PRESERVATION) Cross grain lowers compression and tensile strength but may increase shear strength in some cases. Compression wood, formed in leaning coniferous trees, is weaker than normal wood of the same density (*see* CONIFERALES). Compression failures or localized buckling in wood fibers weaken the wood, particularly in tension under impact forces

Working stresses, used in design of timber structures, are calculated on the basis of standard strength tests, corrected for factors affecting the strength of wood in service. Construction timbers are graded according to their freedom from strength-reducing defects *See* FOREST AND FORESTRY [C.S.]

Bibliography: H D. Tiemann, *Wood Technology*, 3d ed., 1951, U.S Dept. Agric., *Wood Handbook*. Handbook 72, 1955

Wood preservation

The technique of protecting wood through chemical treatment, from fire and from deterioration by the biological agencies of fungi, insects and marine borers Wood preservation skillfully practiced increases the useful life of wood and reduces replacement costs For example, class I railroads through the use of treated ties over a 52-year period (1898–1949) had total approximate savings in material and replacement costs of $6,700,000,000 or a daily saving of $353 000

Deterioration of wood. Properly designed wood structures give long service without special protection, but a large economic loss may result when wood in its natural state is used at high temperatures, in salt water structures, or under climatic conditions that favor the development of harmful fungi and insects (Fig 1)

Fungi Wood-attacking fungi are forms of saprophytic plants Those that inhabit wood live on its organic components and for significant growth require temperatures from 50 to 90°F, a wood moisture content in excess of 25% (based on dry wood), and a supply of oxygen (*see* WOOD DECAY) Unfavorable conditions in one or more of these respects inhibit growth of fungi Wood that is

Fig 1 Costly repairs due to decay may be involved when untreated wood is exposed in places like open porches where water can be absorbed and held by the wood for long periods.

higher radially than tangentially. Increased density, moisture content, and temperature cause higher conductivity in all cases. Below the fiber-saturation point, the average perpendicular-to-grain thermal conductivity K, in cgs units, can be represented at 25°C by the approximate equation

$$K = [60 + \rho(412 + 5.1\,M)] \times 10^{-6}$$

where ρ is density at the moisture content M, at test. The temperature effect is not well known, but conductivity increases about 0.25% per degree C increase.

Specific heat. Although almost constant for different woods, specific heat increases with temperature and moisture content. The specific heat c_o for oven-dry wood is given by

$$c_o = 0.266 + 0.00116\,t$$

where t is degrees C. The specific heat c_m for wet wood at moisture content M is calculated by

$$c_m = (c_o + Mc_a/100)/(1 + M/100)$$

where c_a is the apparent specific heat of the sorbed water which is usually higher than the value of unity given for free water.

Thermal diffusivity. The ratio of conductivity K to the product of density ρ and specific heat c, or $H^2 = K/\rho c$, is defined as thermal diffusivity. It decreases slightly with increasing wood density and moisture content and is almost constant with temperature due to compensating factors.

Thermal expansion. The coefficient of thermal expansion is about 3.5×10^{-6} per degree C along the grain and shows little variation with density. Across the grain, density affects the expansion which is about $50\ G \times 10^{-6}$ radially, and $72\ G \times 10^{-6}$ tangentially, where G is the nominal green specific gravity.

Electrical properties. Dry wood is an excellent insulator, but its electrical characteristics change markedly with increasing moisture content and temperature.

The direct current resistivity decreases almost exponentially with increasing moisture content up to the fiber-saturation point, above which there is relatively little change. Electric moisture meters have been developed which utilize this effect. They must be corrected for temperature since this has a similar effect (Fig. 2). Conductivity is ionic, and activation energies calculated from the strong temperature effect range from about 32 kcal/mole for dry wood down to about 15 kcal/mole at 15% moisture content. *See* ELECTROLYTIC CONDUCTANCE.

The dielectric properties of wood in an alternating current electric field are complex. The dielectric constant and loss factor increase with moisture content and density at almost all frequencies. They are higher when the electric field is parallel rather than perpendicular to the grain, due to the gross structure of the cells as well as to the longitudinal crystallite orientation. Electric moisture meters have been developed to utilize the variation of the dielectric properties with moisture

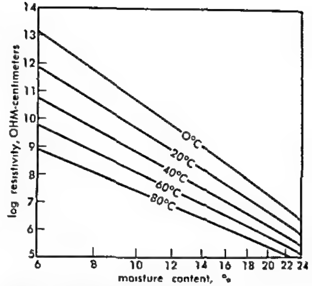


Fig. 2. Relationship of dc electrical resistivity, wood moisture content, and temperature

content. They are affected by variations in density and anatomy of the wood. *See* DIELECTRIC CONSTANT.

Mechanical properties. These are the properties of wood relating to its resistance to deformation by applied forces. Wood is a polymeric material, and therefore its deformation response to external forces is time dependent or rheological (*see* RHEOLOGY). Grain direction, density, moisture content, temperature, and defects also affect the mechanical behavior of wood.

Long term creep and relaxation studies reveal that time constants in the order of 10^9 sec are required to describe the inelastic deformation in wood near room temperatures. Limited experimental data obtained from audiofrequency vibration studies indicate the operation of another unknown mechanism with time constants less than 10^{-4} sec.

Standard mechanical tests on wood must be carried out under specified time rates of strain due to the rheological behavior. These tests are carried out on small clear straight-grained samples of controlled moisture content and temperature. The following relationships are based on such tests.

The parallel-to-grain strength P is usually 10–20 times the strength Q across the grain, due largely to the tubular nature of the wood fibers. At intermediate angles θ, between the grain direction and that of the applied force, certain strength properties N, can be calculated approximately by Hankinson's equation

$$N = PQ/(P \sin^2 \theta + Q \cos^2 \theta)$$

The strength properties in the radial compared with tangential direction vary with the anatomy of the wood.

All strength properties increase with wood density at a given moisture content. However, the relative effect of density on strength Q across the grain is much more pronounced. Stresses across the grain induce bending moments in the cell walls. Increase

hand or principally with special ... chines (see BARK). Conditioning the wood prior to treatment involves either heating or moisture removal. It is essential to proper penetration of preservatives, and is accomplished by procedures such as air seasoning, kiln or tunnel drying, steam-and-vacuum conditioning, heating in the preservative under a vacuum (Boultonizing), and heating in hydrocarbon vapors (vapor drying). Machining before treatment is important, since the cutting or boring of holes in treated wood exposes unpenetrated regions to attack by deteriorating organisms. Improved penetration of preservative in less penetrable species results from incising, a machining procedure for cutting carefully spaced, slitlike holes in surfaces to expose end grain and take advantage of longitudinal preservative penetration.

Fig. 3 Tests on stakes treated with various retentions of chemicals are used as a basis for evaluating wood preservatives.

Wood preservatives. Acceptance of a preservative is generally based on field tests with treated wood stakes (Fig. 3) or posts and on actual service experience involving treated products, such as ties, poles, piling and other structural members. Toxicity to various biological deteriorating organisms is an essential requirement of wood preservatives. Toxicity is determined through biological assays in which measured quantities of the preservative, either in wood blocks (Fig. 4) or in a nutrient substratum, are exposed to the deteriorating organisms. Permanence, another essential, is determined by the vapor pressure, solubility, and other properties of the preservative chemical, and through leaching and weathering tests on the treated wood. Wood preservatives must have favorable penetrating properties, must not corrode metals, and must be safe to handle by those treating the wood and using the finished product. These and other special properties such as paintability, fire retardance, water repellence, gluability, and odor of the treated wood are determined in specially designed laboratory tests.

Preservatives in general use. Wood preservatives in general use are oils, including oilborne and waterborne chemicals.

Oils. These are mostly for outdoor and wet exposure conditions. They do not swell the wood but may contribute to staining and painting difficulties. Coal tar creosote, solutions of creosote with either coal tar or petroleum oil, 5% of pentachlorophenol in petroleum oil, and copper naphthenate (equivalent to 0.75% of copper) in petroleum oil are used for the treatment of products such as ties, posts, poles, piling, and construction timbers, and are covered by specifications of the American government and by standards established by the American Wood Preservers' Association. Only coal tar creosote or creosote–coal tar solutions are acceptable preservatives for the treatment of marine structures. Minimum net retentions of 6–12 lb/ft^3 of wood of these preservatives are recommended in recognized standards for treated products, except for marine installations, which require maximum

Fig. 4 Laboratory soil-block tests are effective for screening preservatives and for indicating the quantity of chemicals needed to inhibit growth of various fungi. Preservative in this case has not been sufficient to prevent growing fungus from attacking treated test block.

dry or submerged in water does not rot. In wood preservation, control is obtained principally through the addition of toxic chemicals to the wood. Mold and stain fungi are confined mostly to sapwood, feed principally on materials stored in the cell cavities, cause objectionable blemishes or discoloration, increase the liquid absorptiveness of the wood, and affect its strength in toughness or shock resistance. Decay fungi feed principally on the cell wall components of cellulose and lignin (*see* CELL WALLS IN PLANTS). Thus they reduce the specific gravity of wood and also seriously affect all strength properties. Brown-rot fungi attack the wood cellulose, whereas both lignin and cellulose are consumed by white rots (*see* CELLULOSE). The heartwood of a few woods contains components that inhibit decay fungi, but the sapwood of all woods is subject to attack. *See* WOOD (ANATOMY AND IDENTIFICATION). So-called "soft" or surface rot and fungus growth found in wood in marine installations are caused by fungi of the Fungi Imperfecti and Ascomycetes groups. *See* ASCOMYCETES; FUNGI IMPERFECTI.

Insects. Some insects damage the appearance of wood or seriously reduce its strength in their efforts to obtain food and shelter. Termites of the ground-inhabiting or subterranean type attack wood structures throughout the United States, but damage is most serious in the southern states (*see* ISOPTERA; TERMITE). Insulation of the wood from the ground to prevent the building of shelter and moisture conduction tubes by the termites, or its impregnation with wood preservatives, are effective control measures. Dry-wood termites cause damage to buildings in the United States principally in southern areas of California and in Florida. Preservative-treated wood resists attack. Paint coatings on wood and screened openings help to prevent attack in buildings, and applications of insecticides and fumigation are effective when usable (*see* FUMIGANT; INSECTICIDE). Bark, ambrosia, and powder-post beetles cause damage to wood in various forms, but other control measures are generally preferable to wood preservation, such as cutting logs during the dormant season (October or November) and allowing them to dry properly to prevent damage caused by bark beetles and ambrosia beetles, or sterilizing the wood with steam at 130–180°F to avoid injury by powder-post beetle. Carpenter ants damage partially rotted wood in buildings and poles (*see* HYMENOPTERA). Measures for decay prevention are also effective against these insects, and insecticides can be applied to existing colonies.

Marine borers. Piling, boats, and other wood structures in salt or brackish water are damaged by borers of the mollusk and crustacean groups which use the wood for shelter and to some extent for food. All woods are subject to attack, but heartwood of a few foreign species has some resistance, such as greenheart, jarrah, azobe, totara, and manbarklak. Wood heavily impregnated with coal-tar creosote or creosote–coal tar solution is highly resistant to mollusk attack and also resists attack by crustacean borers. *See* PELECYPODA.

Fire. Fire resistance in wood structures is obtained by the use of good construction practices that take advantage of the self-insulating and slow-burning properties of wood. Additional fire resistance in hotels, hospitals, ships, schools, military barracks, and public buildings is achieved through impregnation of the wood with fire-retarding chemicals and surface coatings of fire-retardant paints. Effective treatment retards temperature increases during a fire, decreases rate of flame spread, and makes fire more easily controlled. Chemicals used, usually in combinations, are principally ammonium phosphate and sulfate, borax and boric acid, and zinc chloride. Formulations containing zinc chloride combine decay and fire protection properties. About 3 lb of chemical per cubic foot of wood are required for moderate protection and 5–6 lb/ft³ for high protection.

Wood treating characteristics. Wood species vary in the degree to which they can be penetrated with preservatives and in protection resulting from chemical treatment. Sapwood is more easily penetrated than heartwood, and penetration longitudinally exceeds that transversely (*see* XYLEM). Gum and resin deposits and obstructions such as tyloses in the wood elements interfere with liquid movement, as does excessive moisture in the wood. However, some waterborne preservatives penetrate moisture-laden wood readily by diffusion. The heartwood of white oaks, sweetgum, Rocky Mountain Douglas-fir, true firs, and larches is very resistant to treatment with preservatives, whereas the heartwood of maples, birches, ashes, elms, tupelo, ponderosa pine, Coast Douglas-fir (Fig. 2), and western hemlock is either easily penetrated or only moderately difficult to penetrate.

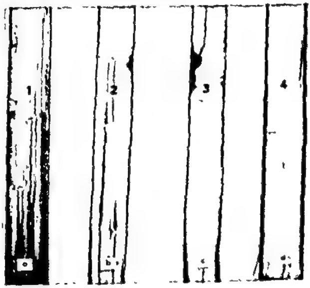


Fig. 2 Round, seasoned posts split after pressure treatment with coal-tar creosote to show penetration of preservative in the heartwood and sapwood of (a) Coast Douglas fir, (b, c) Rocky Mountain Douglas-fir, (d) Western larch.

hand or principally with special ... chines (see BARK). Conditioning the wood prior to treatment involves either heating or moisture re... penetration of pre-

hydrocarbon vapors (vapor drying). Machining before treatment is important, since the cutting or boring of holes in treated wood exposes unpenetrated regions to attack by deteriorating organisms Improved penetration of preservative in less penetrable species results from incising, a machin... spaced slitlike ... ad... tion

Preservatives in general use. Wood preservatives in general use are oils, including oilborne and waterborne chemicals.

Oils These are mostly for outdoor and wet exposure conditions. They do not swell the wood but may contribute to staining and painting difficulties. Coal tar creosote, solutions of creosote with either coal tar or petroleum oil, 5% of pentachlorophenol in petroleum oil, and copper naphthenate (equivalent to 0 75% of copper) in petroleum oil are used for the treatment of products such as ties, posts, poles, piling, and construction timbers, and are covered by specifications of the American government and by standards established by the American Wood-Preservers' Association Only coal tar creosote or creosote–coal tar solutions are acceptable preservatives for the treatment of marine structures Minimum net retentions of 6–12 lb/ft^3 of wood of these preservatives are recommended in recognized standards for treated products, except for marine installations, which require maximum

Fig 3 Tests on stakes treated with various retentions of chemicals are used as a basis for evaluating wood preservatives

Wood preservatives. Acceptance of a preservative is generally based on field tests with treated wood stakes (Fig 3) or posts and on actual service experience involving treated products, such as ties, poles, piling, and other structural members Toxicity to various biological deteriorating organisms is an essential requirement of wood preservatives Toxicity is determined through biological assays in which measured quantities of the preservative, either in wood blocks (Fig 4) or in a nutrient substratum, are exposed to the deteriorating organ-

leaching and weathering tests on the treated wood Wood preservatives must have favorable penetrating properties, must not corrode metals, and must be safe to handle by those treating the wood and using the finished product. These and other special properties such as paintability, fire retardance, water repellence, gluability, and odor of the treated wood are determined in specially designed laboratory tests

Fig 4 Laboratory soil-block tests are effective fo[r] screening preservatives and for indicating the quantit[y] of chemicals needed to inhibit growth of various fung[i] Preservative in this case has not been sufficient to pr[e]vent growing fungus from attacking treated test bloc[k]

possible retentions. Creosotes from wood tar, water-gas tar, and oil tar are less effective than coal-tar creosote and have limited use.

Solutions of pentachlorophenol and copper naphthenate and other oil-borne chemicals such as zinc naphthenate, copper-8-quinolinolate, tetrachlorophenol, and chloro-2-phenylphenol, alone or in combination, are often sold under proprietary (trade) names for special-purpose treatments of wood. These may contain water repellents when used for such items as millwork. Solvents used with oil-borne preservatives vary from heavy fuel oils to mineral spirits, depending on use requirements.

Water-borne. Water-borne preservatives are usually lower in cost than oils and less likely to involve odor and painting problems. They are used principally for less severe exposure conditions, but several are highly resistant to leaching and appear to perform as well as preservative oils in contact with the soil or fresh water. Water in the treating solutions causes swelling of the wood, however, and redrying to levels meeting use requirements is necessary. Treatment specifications usually require retentions of less than 1 lb of water-borne preservative per cubic foot, so the weight of the wood after redrying is not significantly increased. Preservatives of this type contain compounds of copper, zinc, fluorine, and arsenic, sometimes in combination with chromates. Some are covered by patents and are sold under proprietary names. Various American government specifications and the Standard P-5 of the American Wood-Preservers' Association define the composition of recognized water-borne preservatives.

Application of preservatives. Some wood products, such as window sash, are exposed to a very limited decay hazard and require less protection than construction timbers, poles, crossties, and piling, which are required to give long service under severe conditions. In general, however, the extent of protection furnished by a preservative is determined by the effectiveness of the method of application and by the resulting penetration and retention of the chemicals. Pressure impregnation is the most common and generally the most effective method of treatment, although there are several nonpressure applications of moderate to good effectiveness and varying simplicity.

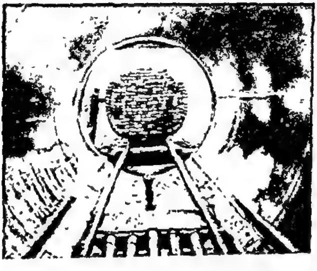


Fig. 5. A tramload of lumber removed from pressure-impregnating cylinder after treatment with a water-borne preservative.

Pressure treatments. Wood is placed in horizontal steel cylinders 4–9½ ft in diameter and up to 180 ft in length and impregnated with preservatives (Fig. 5) under pressure of 50–200 psi. In the full-cell process, used for applying water-borne chemicals and for treatment with creosote for marine uses, an initial vacuum is used prior to impregnation with the aim of getting maximum retention and penetration of preservative. In the empty-cell process, initial air at various pressures is used prior to impregnation to control preservative retentions and to provide maximum penetration. A final vacuum is commonly used after impregnation by the empty-cell process and may be used after full-cell treatment to remove surplus preservative from the surface of the wood. Empty-cell treatment is used for lumber, posts, and poles where oil preservative retentions of 4–16 lb/ft³ are specified.

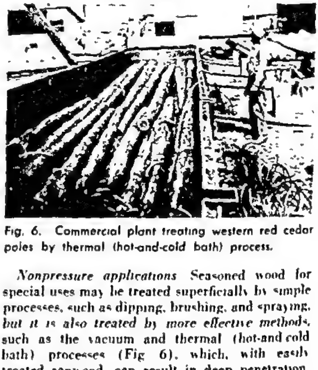


Fig. 6. Commercial plant treating western red cedar poles by thermal (hot-and-cold bath) process.

Nonpressure applications. Seasoned wood for special uses may be treated superficially by simple processes, such as dipping, brushing, and spraying, but it is also treated by more effective methods, such as the vacuum and thermal (hot-and-cold bath) processes (Fig. 6), which, with easily treated sapwood, can result in deep penetration. Several nonpressure processes in limited use for unseasoned wood depend on diffusion of water-borne chemicals into the water in the wood. One is the double diffusion process, which forms effective preservatives of low solubility in the wood by combining two separately applied soluble compounds such as copper sulfate and sodium arsenate. Various nonpressure applications of preservatives are also used in pole maintenance at the groundline of standing poles. Preservatives have been applied by various methods, with limited success, to living or freshly felled trees, including unpeeled posts and poles. See FOREST AND FORESTRY. [J.O.B.]

Bibliography: J. O. Blew, Jr., et al., *Evaluating Wood Preservatives*, Proc. Forest Products Re-

search Soc., vol. 5, 1951; J. O. Blew, Jr., et al., *Symposium on Conditioning Wood Prior to Preservative Treatment*, Proc. Am. Wood-Preservers' Assoc., vol. 54, 1958; Forest Products Lab., *Wood Handbook*, US Dept. Agr. Handbook 72, 1955,

Agr Handbook 40, 1952.

Woodchuck

A large, heavy-bodied member, *Marmota monax*, of the family Sciuridae, up to 30 in long, with a short tail, small ears, strong front legs, and heavy claws. It is more commonly called ground hog and is sometimes known as the whistle pig because of its habit of whistling when alarmed. It can be destructive to gardens and forage but is usually of little importance. It is sometimes used for food.

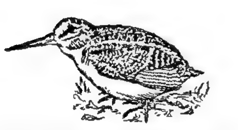
Woodchucks dig large dens in which they rear their young and hibernate during the winter. They eat coarse, woody plant food, as well as some animal food. They are usually gray to black in color, with some reddish marks.

There are three western species of *Marmota*, all called marmots, and all quite similar to the woodchuck of the eastern United States. Other marmots occur in Mexico, Asia, and Europe. The burrows of all of them are used by other animals, notably rabbits. Marmots provide a limited amount of sport shooting. See RODENTIA [J D B.]

Woodcock

A woodland member, *Philohela minor*, of the family ...olopacidae. It is short-legged and short-tailed, ...th a chunky body and a long bill. The woodcock ...streaked above with varying shades of black, ...ay, buff, and brown, providing it with an effective ...mouflage for its woodland habitat. It frequents ...mp thickets and damp open spots in the forest ...here it probes the mud for earthworms and in...ct grubs.

The woodcock may be found in the deciduous ...rests of eastern North America but is rare except ...ally; it winters in numbers in Louisiana. It is ...nsidered a game bird and provides some shoot...ng in a few localities. See CHARADRIIFORMES [J D B]



The woodcock, *Philohela minor*, length to 1 ft. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Woodpecker

Any of over 200 species of birds of the family Picidae, order Piciformes, a virtually cosmopolitan family (not in Australia) modified for climbing tree trunks and drilling in wood. The woodpecker has a strong bill; long and extensile tongue; short, stiff tail; and hard skull. Most woodpeckers drill in wood for insect grubs which they use as their food. All of them drill nests in tree trunks. Some, like the flickers, *Colaptes* sp., spend much time on the ground feeding upon ants and other terrestrial insects. Most woodpeckers are beneficial, destroying wood-boring insects not attacked by other predators. The damage they do to trees is more imaginary than real. The sapsuckers, *Sphyrapicus* sp., which drill rows of holes in maples and other trees with free-flowing sap, frequently damage trees by girdling them, but are not common enough to be a serious problem.

The redheaded woodpecker, *Melanerpes erythrocephalus*, length to 9¾ in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Largest and rarest of the 21 species of woodpeckers in the United States is the ivory-billed woodpecker, *Campephilus principalis*, last reported from Florida in 1950, and possibly extinct. The pileated woodpecker, *Dryocopus pileatus*, is a similar species, almost as large, still present in the forests of the United States and Canada. The redheaded woodpecker, *Melanerpes erythrocephalus*, is a pest of telephone and power lines because of its habit of drilling holes in utility poles.

Other familiar species include the small downy woodpecker, its slightly larger duplicate, the hairy woodpecker, the red bellied woodpecker; and the acorn-storing California woodpecker. See PICIFORMES [J.D B.]

Woodworking

The shaping and assembling of wood and wood products into finished articles such as mold patterns, furniture, window sashes and frames, and boats. The pronounced grain of wood requires modifications in the working techniques when cutting with the grain and when cutting across it. Five principal woodworking operations are sawing, planing, steam bending, gluing, and finishing. To shape round pieces, wood is worked on a lathe. *See* TURNING (WOODWORKING).

Sawing. Wood is sawed by cutting or splitting its fibers by the continuous action of a series of uniformly spaced teeth alternately staggered to move in closely parallel work planes. Action of the cutting teeth produces a path or kerf of uniform width through the workpiece from which the fibers have been severed and removed. Sawing across the grain or cell structure of the wood is called crosscutting. Cutting parallel with the grain of the piece is referred to as ripping. Saw teeth are bent alternately to the left and right to provide clearance for the blade. Some blades include straight raker teeth for cleaning fibers from the cut.

Wood cutting saws may be classed as either handsaws or power operated saws. Either group consists of numerous types and designs.

Handsaws. Each type handsaw is designed to accomplish one specific type of sawing operation most effectively.

Crosscut handsaws are made with about 8 teeth per inch of length. Ripsaws for cutting with the grain usually have about 5½ teeth/in Fine tooth saws for finishing or cabinet work may have as many as 10–16 teeth/in. of length

A backsaw is a fine-tooth saw with its upper edge stiffened to ensure straight cuts. Keyhole saws with their narrow, tapered blades are used for cutout work where sharp turns are required A compass saw has a handle with several attachable blades of varying widths, making it suitable for a variety of work Coping saws have narrow blades usually about ¼ in. wide. The blade is held taut in a frame which is equipped with a handle. The narrow blade and high backed frame make the saw suitable for shaping or cutout work.

Power saws. Power operated woodworking saws are usually combined with auxiliary equipment that enables them to perform various sawing operations.

Bench or circular saws are the common woodworking type of power saw (Fig. 1). Depending on construction, either power arbor or worktable may be raised, lowered, or tilted. The saw may be used for crosscutting, ripping, or resawing, and for beveled as well as tapered cuts. Molding cutters along with those designed for making rabbet, tenon, and dado joints are also used. Accessories for sanding, buffing, and polishing are available for most models.

Band saws are basically a flexible band of steel running over two vertical pulleys (Fig. 2). The band or blade has teeth on one side and is operated under tension. The wide distance or throat between the cutting portion of the blade and the rear blade guide and support arm adapt the band saw for cutout work or sawing on large flat pieces.

Scroll saws are used for work similar to that performed by the band saw. The continuous band type of blade is replaced by a short, vertically reciprocating blade (Fig. 3).

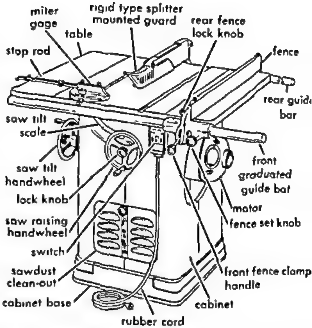


Fig. 1. Bench circular saw with tilting arbor is used for parting or slotting and can make cuts as long as working space permits. (*Delta*)

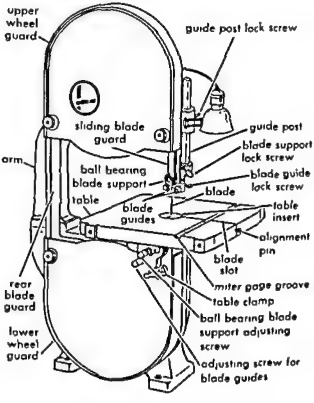


Fig 2. Band saw, being narrow, can make curved as well as straight cuts even in thick pieces. (*Delta*)

Radial saws have their circular blade mounted above the worktable The blade and motor are suspended from an overarm that allows travel across the workpiece; a pivot permits cuts to be taken at any angle (Fig 4). Usually the saw may be raised and lowered as well as tilted at an angle. Crosscutting, ripping, mitering, and beveling may be performed, accessories and attachments permit other circular cutting tool operations such as dadoing, molding, drilling, and sanding.

Portable handsaws consist of a circular blade and electric motor plus the necessary frame, handles, baseplate, and guards An electric cord of reasonable length permits the saw to be manually moved and positioned for the desired cut. Some models may be fastened to a special frame or table.

Portable saber saws are compact units consisting of an electric motor, a straight saw blade with reciprocating mechanism, handle, baseplate, electric cord, and other necessary parts The saw's lightness and its narrow blade fastened at only one end make it adaptable to many types of cutting including cutout work and shaping.

Planing. Flat or uniformly contoured surfaces of wood are roughed down, smoothed, or made level by the shaving and cutting action of a wide-edged blade or blades Planing may be accomplished either manually or by power operated tools

Hand planes. Manually operated planes are classified as either bench or block types A bench plane is used for shaving with the grain of the wood, whereas a block plane is designed for cutting across the grain A block plane is usually small; bench planes vary in size and type. The common bench types are the smoothing, jack, fore, and jointer.

The *smoothing plane,* 5.5–10 in long, is best suited for smoothing small areas The somewhat larger jack plane may be used for roughing down or leveling Fore and jointer planes are still larger in size, with the latter being approximately 22–24 in long It is used to plane long surfaces.

Special hand planes are the rabbet plane used to cut recesses for rabbet joints, the modelmaker's plane with its blade and sole curved so that it can remove excess wood from a curved surface, the scrub or roughing plane which has heavy, rounded blades making it suitable for cleaning up rough boards; and the circular plane with a flexible steel bottom that may be adjusted to fit a curved surface.

Power planes Power operated planes vary in size and design according to the application and type of work handled The planer usually found in the woodworking shop is the *jointer,* frequently called a jointer planer (Fig 5).

The jointer is designed with its cutting blades or knives fastened in a rotating cutterhead The length of the knives classify the machine as to the width of board that it can surface.

Two tables, one in front of the cutterhead and one behind it, support the workpiece as it is pushed through the path of the knives The front table is

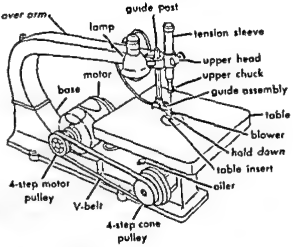


Fig 3. Scroll saw cuts sharp turns in thin pieces (*Delta*)

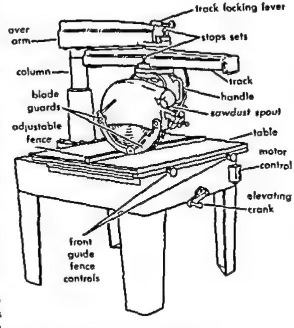


Fig 4 Radial saw allows wood to be clamped in position while it is being cut (*Delta*)

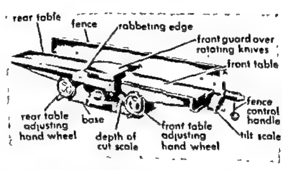


Fig 5 Six-inch jointer planer is used for surfacing, rabbeting, beveling, tapering, molding, and cutting round tenons (Delta)

set lower than the highest point on the arc of the rotating knives by the amount to be planed from the board. The rear table must be aligned exactly with the high point of the knives so that the workpiece will not pass through the cutting path at an angle.

A tilting fence or guide is provided along one side of the tables. The opposite side of the front table has a ledge which is used to support the workpiece for rabbeting cuts.

Steam bending. Wooden members are bent or formed to a desired shape by pressure after they have been softened or plasticized by heat and moisture. If thick pieces of wood are to be bent to a permanent shape without breaking, some form of softening or plasticizing such as steaming is necessary.

When a piece of wood is bent, its outer or convex side is actually stretched in tension while its concave side is simultaneously compressed. Actually, plasticized wood can be stretched but little. It can, however, be compressed a considerable amount. When a piece of plasticized wood is successfully bent, the deformation is chiefly compression distributed almost uniformly over the curved portion. Curvature results from many minute folds, wrinkles, and slippages in the compressed area.

Steaming. Although soaking wood in water softens it somewhat, a combination of heat and moisture can produce a degree of plasticity approximately ten times that of dry wood at normal temperatures. Wood need not be steamed to its maximum plasticity in order to be bent. Wood steamed at atmospheric pressure bends in most cases as well as wood steamed at higher pressures. Higher pressures also tend to overplasticize the wood; this results in an increased number of bending failures. Treatment of wood with boiling water has approximately the same effect as saturating it with steam at atmospheric pressure. The boiling water treatment is usually employed only when a portion of a wooden piece requires softening.

Dry wood which has a moisture content of 12% or less must have its moisture increased to approximately 15% to make it suitable for moderate bends. If more severe bends are required, additional moisture must be added to the surface areas. Wood which already has a moisture content of 20-25% needs no further saturation, even for severe bending.

The time required for steaming is directly related to the amount of moisture already present in the wood. In general, dry stock is steamed 1 hour per inch of thickness and green stock, ½ hour per inch of thickness. Steaming is usually done in a closed retort suitable for zero or low gage pressure steam. Steam should enter the retort or steam box through water standing in the bottom in order that steam in the box will be wet or saturated.

Wood bending methods may be classed as made without end pressure (free bends) or made with end pressure. On thick pieces, only slight curvatures are feasible by free bending. Bending with end pressure is necessary to obtain the required compression and to prevent tensile failures in moderate or severe bends. The most common method of bending with end pressure is by means of a metal strap fitted with end blocks or clamps. The strap is placed against the convex side of the piece, and bending pressure is applied at the end of the strap by some suitable means (Fig. 6). The end blocks must be applying end pressure simultaneously on the wood to prevent tensile stress and also to supply the necessary compression.

Other devices used for bending plasticized wood are variable position rollers, hot-plate form presses, and special mechanical devices.

When a workpiece is removed from the bending device, it has a tendency to spring back somewhat. This is counteracted by holding the piece in position until it has dried or set. Frequently springback is compensated for by overbending.

The selection of a species of wood to use as a moderately bent member is governed primarily by its suitability and availability. However, if a severe bend is required, the wood must be selected chiefly for its bending qualities. In general, hardwoods are of better bending quality than softwoods. Pieces should be of fairly straight grain and free of knots or other defects. White and red oak, hickory, elm, ash, beech, birch, maple, mahogany, walnut, and sweet gum are species commonly used for bending.

Gluing. Wood pieces may be fastened together by the adhesive qualities of a substance that sets or hardens into a permanent bond. Certain glues require heat and pressure to complete the process.

Animal, vegetable, liquid, casein, blood albumen, and synthetic resin are the principal classes of woodworking glues. Of these, casein, which has a skim milk base, is probably the most popular and practical for the small workshop. Animal and liquid glues are used to some extent by cabinet makers. The remaining types require special equipment such as heating facilities and presses, which tend to limit them to commercial use. Some thermo-

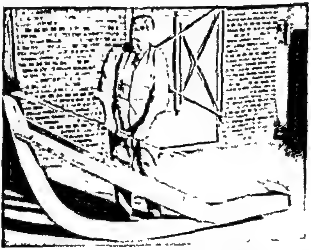


Fig. 6. Bent boat rib prepared for drying with end pressure applied by strap and tie rods. Wood stays further hold work in position. (Forest Products Laboratory, Madison, Wisconsin)

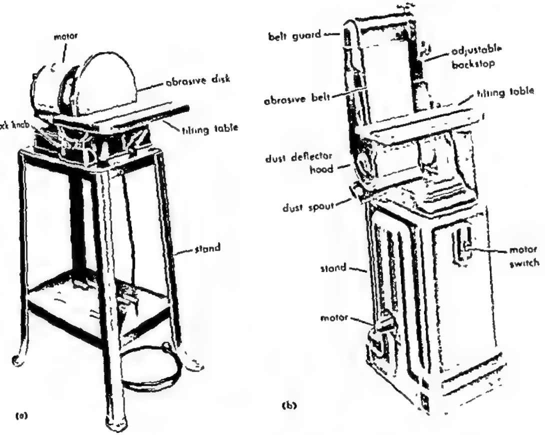


Fig 7. Types of power sander (a) Disk. (b) Belt (Delta)

setting resins require temperatures of over 300°F and pressures as high as 250 psi depending on the wood *See* GLUE.

Casein glue comes in the form of a powder and is mixed with cold water It may be used in a warm or cold room, and no heating of glue, water, or wood is required The glue is rather thick when mixed and should be spread evenly over the piece

Prior to any gluing, the mating surfaces should be clean and dry and fit closely After a thin coat has been applied to both surfaces, they should be allowed to set until the glue is nearly dry A second coat is permitted to stand until it becomes tacky to the touch The pieces are fitted together

Clamps may be used to hold pieces together However, clamps should not be so tight that pieces are misaligned or that much glue is squeezed out at the joint. Clamps are usually left on the work about one day.

Finishing. The preparation and sealing or covering of a surface with a suitable substance in order to preserve it or to give it a desired appearance is the finishing operation

The preparation and conditioning of a surface may include cleaning, sanding, use of steel wool, removing or covering nails and screws, gluing or fastening loose pieces, filling cracks and holes with putty or crack filler, shellacking, and dusting An inconsequential item with a painted surface does not require the thorough surface preparation that is appropriate for a piece of fine furniture The quality of surface conditioning directly affects the end result

Surface preparation Wooden surfaces to be painted are cleaned and sanded Sanding is done with the grain of the wood, using a sandpaper block to keep the surface even, or a power sander (Fig 7). Use of steel wool instead of sandpaper in corners and on rounded surfaces is advisable. Nails and screws should be removed or sunk below the surface and these holes, as well as any cracks, should be filled with putty or crack filler New wood is covered with a coating of thinned orange shellac before applying paint. The undercoat tends to prevent the paint from drying unevenly, and thus producing flat or glossy spots over knots or other irregularities in the wood.

If the surface is to be varnished, stained, or fin-

should be tinted to match the wood After the filler material has hardened, the entire surface should be sanded with coarse sandpaper. When the rough spots have been removed, sanding should be finished with a no. 1 sandpaper On fine hardwood surfaces or on furniture from which the varnish has been removed, only fine grades of sandpaper run-

set lower than the highest point on the arc of the rotating knives by the amount to be planed from the board. The rear table must be aligned exactly with the high point of the knives so that the workpiece will not pass through the cutting path at an angle.

A tilting fence or guide is provided along one side of the tables. The opposite side of the front table has a ledge which is used to support the workpiece for rabbeting cuts.

Steam bending. Wooden members are bent or formed to a desired shape by pressure after they have been softened or plasticized by heat and moisture If thick pieces of wood are to be bent to a permanent shape without breaking, some form of softening or plasticizing such as steaming is necessary.

When a piece of wood is bent, its outer or convex side is actually stretched in tension while its concave side is simultaneously compressed. Actually, plasticized wood can be stretched but little. It can, however, be compressed a considerable amount. When a piece of plasticized wood is successfully bent, the deformation is chiefly compression distributed almost uniformly over the curved portion Curvature results from many minute folds, wrinkles, and slippages in the compressed area.

Steaming. Although soaking wood in water softens it somewhat, a combination of heat and moisture can produce a degree of plasticity approximately ten times that of dry wood at normal temperatures Wood need not be steamed to its maximum plasticity in order to be bent. Wood steamed at atmospheric pressure bends in most cases as well as wood steamed at higher pressures Higher pressures also tend to overplasticize the wood; this results in an increased number of bending failures Treatment of wood with boiling water has approximately the same effect as saturating it with steam at atmospheric pressure The boiling water treatment is usually employed only when a portion of a wooden piece requires softening.

Dry wood which has a moisture content of 12% or less must have its moisture increased to approximately 15% to make it suitable for moderate bends If more severe bends are required, additional moisture must be added to the surface areas Wood which already has a moisture content of 20-25% needs no further saturation, even for severe bending.

The time required for steaming is directly related to the amount of moisture already present in the wood. In general, dry stock is steamed 1 hour per inch of thickness and green stock, ½ hour per inch of thickness. Steaming is usually done in a closed retort suitable for zero or low gage pressure steam. Steam should enter the retort or steam box through water standing in the bottom in order that steam in the box will be wet or saturated

Wood bending methods may be classed as made without end pressure (free bends) or made with *end pressure*. On thick pieces, only slight curvatures are feasible by free bending. Bending with end pressure is necessary to obtain the required compression and to prevent tensile failures in moderate or severe bends. The most common method of bending with end pressure is by means of a metal strap fitted with end blocks or clamps. The strap is placed against the convex side of the piece, and bending pressure is applied at the end of the strap by some suitable means (Fig. 6). The end blocks must be applying end pressure simultaneously on the wood to prevent tensile stress and also to supply the necessary compression.

Other devices used for bending plasticized wood are variable position rollers, hot-plate form presses, and special mechanical devices

When a workpiece is removed from the bending device, it has a tendency to spring back somewhat. This is counteracted by holding the piece in position until it has dried or set. Frequently springback is compensated for by overbending

The selection of a species of wood to use as a moderately bent member is governed primarily by its suitability and availability. However, if a severe bend is required, the wood must be selected chiefly for its bending qualities. In general, hardwoods are of better bending quality than softwoods Pieces should be of fairly straight grain and free of knots or other defects White and red oak, hickory, elm, ash, beech, birch, maple, mahogany, walnut, and sweet gum are species commonly used for bending

Gluing. Wood pieces may be fastened together by the adhesive qualities of a substance that sets or hardens into a permanent bond Certain glues require heat and pressure to complete the process.

Animal, vegetable, liquid, casein, blood albumen, and synthetic resin are the principal classes of woodworking glues. Of these, casein, which has a skim milk base, is probably the most popular and practical for the small workshop Animal and liquid glues are used to some extent by cabinet makers The remaining types require special equipment such as heating facilities and presses, which tend to limit them to commercial use Some thermo-

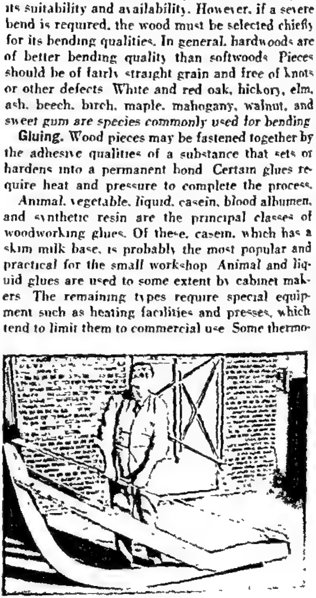


Fig. 6 Bent boat rib prepared for drying with end pressure applied by strap and tie rods Wood stays further hold work in position (*Forest Products Laboratory, Madison, Wisconsin*)

Table 1. Shorn wool in United States, 1956 and 1957*

Year	Number sheep shorn†	Average weight per fleece, lb	Total production, lb	Average price per pound, cents	Total value
1956	28,502,000	8 37	238,569,000	44 2	$105,511,000
1957	28,508,000	8 26	235,366,000	54 4	128,032,000

* From USDA Agricultural Marketing Service Statist. Bull. 230, July, 1958.
† Includes shearing at commercial feeding yards.

ceutical and cosmetic industries for lanolin compounds because it can be absorbed by the human skin. Both glands open into the follicle. Their secretions function as lubricants and protectants for the wool fiber as it grows.

Growth. Blood capillaries in the papilla nourish the growing fiber, which consists of two parts, the root and the shaft. The root, or living part, is beneath the skin surface. The shaft, which protrudes from the mouth of the follicle, is dead. Physical and chemical differences between the root and shaft are listed in Table 2

Table 2. Physical and chemical differences between root and shaft of wool fiber*

Root	Shaft
Soft and easily crushed	Tough and horny
Cells roundish	Cells elongated
Positive test for nucleic acid	Negative test for nucleic acid
Nuclei stained with hematoxylin	Nuclei unstained with hematoxylin
Cytoplasm granular in appearance	Cells distinctly fibrous
Not birefringent	Birefringent
Positive test for sulfhydryl groups	Negative test for sulfhydryl groups
No Allwoerden reaction with chlorine water	Many large Allwoerden "sacs"

* From J M Matthews and H R Mauersberger, *Textile Fibers*, 6th ed , Wiley, 1954

A cross section of the shaft (Fig 3) reveals three layers the cuticle, the cortex, and the medulla The last normally is absent in fine wools and infrequent in ...

The cuti...
and has ov...
which poi...

The cortical layer lies beneath the cuticle Its long spindle cells contribute tensile strength and elasticity to the wool fiber The cortex is further divided bilaterally into a para- and an orthocortex This is demonstrated by the use of a preferential dye (acid or basic), which is taken up by the orthocortex, and may be seen in a cross-sectional view.

In medium and coarse wools, the third layer, the medulla, is comprised of superimposed, honeycomb-like cells filled with air Medullation is a problem to the manufacturer inasmuch as fibers possessing it have lower spinning properties, are lustrous, straight, and coarse In piece-dyed fabrics, they produce a skittery effect if they are dyed a lighter shade

Physical properties. Wool's major physical characteristics include fiber diameter (fineness or

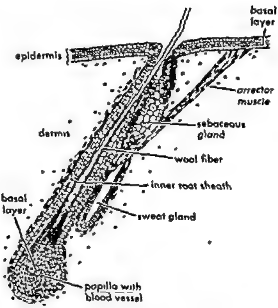


Fig 2 Longitudinal section of a completely developed nonmedullated wool follicle magnified about X100 (After Wildman)

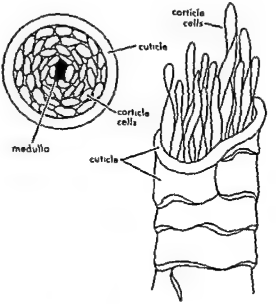


Fig 3 Cross section of wool fiber showing cuticle, cortex, and medulla

ning from no. 0 down to no. 000 should be used. Rubbing must be with the grain, and grit and dust must be removed frequently. If paint remover has been used, the surfaces should be thoroughly cleaned with turpentine and then washed with hot water and soap to remove any wax.

Finish application. Methods used to apply finish coverings to wooden surfaces vary with the substance. In any case, the surface should be wiped clear of dust just prior to applying the material The atmosphere and equipment used should be as dust-free as possible.

Paint may be applied by brush or spray as the situation permits. The priming coat and each additional one should be thoroughly dry before applying the next one. On woodwork or painted furniture after each coat prior to the final one, the surface should be sanded lightly to remove any dust or particles stuck to the paint. If a brush is used, the paint should be applied with long even strokes. A thin coat will go on more evenly and tend to prevent the paint from running or forming lumpy spots at corners and edges.

Stains penetrate the pores of wood but do not fill or close them. Oil stains are the ones most commonly used except on furniture, where water stains are the rule. A few moments after a stain has been applied to a wood surface, any surplus should be wiped off with a lint-free cloth. The time that a stain remains on the wood before being wiped off determines the resulting shade. Although shellac or varnish may be applied over a stain, usually a liquid or paste filler is used to seal the pores in the wood. If varnish is used, two or three coats are generally required.

Varnishes should be applied to a surface quickly and liberally. After most of the varnish in the brush has been applied on an area, it should be spread out evenly. Strokes should be with the grain, and all marks or laps should be brushed out with a nearly dry brush. Varnish used on the first coat should be thinned with about a quart of turpentine to a gallon of varnish.

Each coat of varnish should be allowed to dry for several days before applying the next. Sanding should be done with no. 000 sandpaper. Thorough removal of dust is a necessity. Varnishing should only be attempted at temperatures of 65°F or above.

Shellac, which is thinner than varnish, dries quickly and must be applied to a surface rapidly. All parts of the surface should be carefully covered with the shellac. Several hours should be allowed for drying between coats. White shellac is used for light woods with a natural finish, while orange shellac is used on all other types of wood. In general, shellac should be used with the same procedure that applies to varnish.

Linseed oil or paste wax are frequently used in finishing furniture. Rubbing may include the use of pumice and alcohol. *See* WOOD FINISHING; WOOD PRESERVATION; *see also* WOOD FIBER PRODUCTS; WOOD PHYSICS [A.T.]

Wool

Wool provides warmth and physical comfort that cotton and linen fabrics cannot give. These qualities, combined with its soft resiliency, make wool a necessity for apparel and for rugs and blankets (Fig. 1).

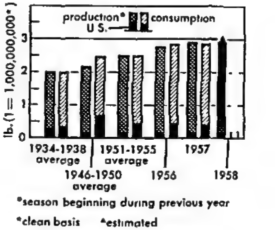


Fig 1. World production and consumption of wool. (USDA, AMS)

Sheep are generally shorn of their fleeces in the spring, but the time of shearing varies in different parts of the world (Table 1). Sheep are not washed before shearing. Sometimes they are dipped into an antiseptic bath, but only when prescribed by law. Formerly, sheep were shorn by hand, but today the fleeces are usually removed in one piece by machine clippers, which shear closer and faster than hand clippers. Wool shorn from young lambs differs in quality from that of older sheep, and wool from live sheep is different from that of dead sheep.

Domestic wool reaches the mill in loosely packed bags, imported wool comes in tightly compressed bales. Each fleece contains different grades, or sorts, of wool, and the raw stock must be carefully graded and segregated according to length, diameter, and quality of fiber. Wool from different parts of the body of the sheep differs greatly. The shoulders and sides generally yield the best quality of wool, because the fibers from those parts are longer, softer, and finer.

Wool technology is that branch of animal science concerned with investigating the structure, growth characteristics, and chemical properties of wool affecting and determining commercial use. The term wool covers the fibers of sheep, Angora goats, camels, alpacas, llamas, and vicunas. In this article, however, wool technology refers to sheep fiber only.

Fiber structure. Wool is epidermal in origin, a complex, organized structure growing from a follicle buried in the dermis of the skin (Fig. 2) Associated with the follicle are two glands, the sebaceous and the sudiferous. The sebaceous gland secretes wool grease, which, when refined, is lanolin. The sudiferous gland secretes sweat, or suint. Collectively, grease and suint in the raw fleece are called yolk which is widely used in the pharma-

Table 4. Commercial lengths and grades of wool

American system	Fine	½ Blood	⅜ Blood	¼ Blood	Low ¼ blood	Common	Braid
Spinning count	80s–70s, 64s	62s–60s	58s–56s	54s–50s	48s–46s	44s	40s–36s
Commercial length classes	Staple length by grade in inches*						
Staple	2.5 and longer	3.0 and longer	3.5 and longer	4.0 and longer	4.5 and longer	5.0 and longer	5.5 and longer
Good French combing	2.0	2.5	3.0	3.5			
Average French combing	1.5	2.0	2.0	2.5			
Short French combing	1.0	1.5					
Clothing and stubby	Under 1.0	Under 1.5	Under 2.0	Under 2.5	Under 4.5	Under 5.0	Under 5.5

*The length designations are based on unstretched staple length and represent a minimum length for the bulk of the staples in a sample.

Fig. 7 Micronaire, a device for measuring fiber diameter in microns

ple classification also increases for each successively coarser grade.

Clean wool yield. Extraneous matter in grease wool is extremely variable in quantity. Such matter consists of grease, suint, dirt, vegetable matter, and moisture. All of these except moisture are removed in the scouring (washing) process. This percentage loss in weight may vary from 32–78%, and is called shrinkage. The clean scoured wool remaining is called the yield.

Because wool is unmanageable after washing, the fiber is dipped in, or sprayed with, a light emulsion of olive or mineral oil to prevent it from becoming brittle and to lubricate it for the spinning operation. If the wool is to be dyed in the raw stock, it is dyed at this stage.

The percentage of extraneous matter in wool is a major factor in determining its market value. Estimates of the yield are determined either by laboratory testing of samples, or through visual appraisal by the buyer. Because of the variability in shrinkage according to length and grade, visual appraisal of clean wool yield may be inaccurate. The laboratory method known as the core test is more reliable and consistent.

Core test. By means of a motor-driven pressure tube, a core sample of wool is withdrawn from a bag or bale (Fig. 8). Core-sampling patterns and the number of cores taken are prescribed according to the number of bags or bales in the lot. Each sample is then tested by standard procedures for moisture, ash, vegetable matter, and grease content, and is then adjusted to the standard condition of 14% impurities or 12.0% moisture, 0.5% ash, and 1.5% residual grease.

Chemical characteristics. Wool is primarily a protein, keratin (Table 5). ...

...istics of the fiber as a whole. Such reagents are oxidizers, reducers, alkalies, and light. The amphoteric nature of wool is a ready-made tool for the dyer and colorist.

Action of halogens. Treatment of the fiber with the halogens leads to absorption and chemical change. Chlorination causes wool to become yellow, harsh, and lustrous, and to lose its felting characteristics, with a corresponding rise in rate of dye absorption.

Action of heat. If briefly heated in dry air at 100–105°F, wool becomes harsh and loses strength and moisture. Normal moisture, softness, and strength is regained upon return to moist, cool conditions. Wool decomposes over extended periods of heat treatment.

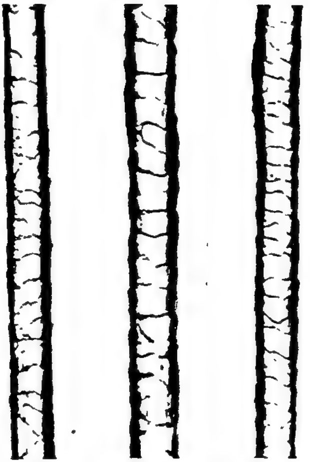


Fig. 4. Scale formation of Delaine Merino wool fibers

grade), staple length, and clean wool yield. Also significant are soundness, color, luster, and content of vegetable matter. Grade refers specifically to mean fiber diameter and its variability. Fiber diameter is the most important manufacturing characteristic. Fleeces are commercially graded visually through observation and handling by men of long experience in the industry. Degree of crimp and relative softness of the fleece are important deciding factors employed by the graders (Fig. 5).

Grading systems. Two systems of grading fleeces are practiced, the American and the spinning count. The seven grades of the American system are inadequate for modern manufacturing systems. The spinning-count system (Table 3) provides for fourteen grades, the range in fiber diameter for the various spinning counts proposed by the American Society for Testing Materials (ASTM).

Measuring devices. For meticulous laboratory research, fiber diameter is expressed in microns (μ) according to methods set forth by the ASTM. Devices utilized in such work include the Hardy sectioning device (Fig. 6) and the micronaire (Fig. 7).

Advantages of grading. The accurate grading of wool is important to producer and manufacturer for the following reasons: (1) it is mechanically impossible to make soft, full-handling worsted fabric from coarse wool; (2) manufacturers must use graded wools for securing desired effects in finished goods; (3) textile machinery either is designed to handle certain grades exclusively, or must be adjusted for each grade; (4) graded wools gain a market advantage; (5) ranch-graded wools give the producer an index of the variability of his flock so that corrective measures can be applied in selecting breeding sheep; and (6) shrinkage estimates are more accurate on graded lines of wool than on ungraded lines.

Length. While wool is being graded, it is classified by length into three major categories: staple, French combing, and clothing (Table 4). Usually longer wools within grade are more valuable than shorter wools because they shrink less and are less wasteful in manufacture. As fiber diameter increases, staple length increases. The length for sta-

Table 3. Proposed specifications for grade or fineness of wool (ASTM)*

Grade	Fineness range, μ (average diam.) min	max	Grade	Fineness range, μ (average diam.) min	max
80s	17.7	19.1	54s	27.9	29.3
70s	19.2	20.5	50s	29.4	30.9
64s	20.6	22.0	48s	31.0	32.6
62s	22.1	23.4	46s	32.7	34.3
60s	23.5	24.9	44s	34.4	36.1
58s	25.0	26.4	40s	36.2	38.0
56s	26.5	27.8	36s	38.1	40.2

* Numerical terms for grade are used internationally and represent the maximum spinning capacity of wool of that fineness.

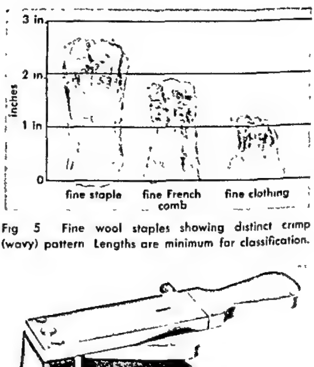


Fig. 5. Fine wool staples showing distinct crimp (wavy) pattern. Lengths are minimum for classification.

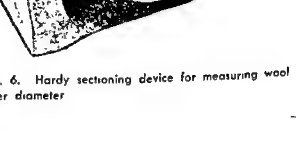


Fig. 6. Hardy sectioning device for measuring wool fiber diameter

To correct wrong impressions concerning the use of remanufactured wool, and also to protect consumers against unscrupulous practices, the United States government passed the Wool Products Labeling Act. This provides that every article of wool clothing must be labeled according to the type of wool used in its manufacture. The label must state (1) amount of wool fiber in the fabric; (2) percentage by weight of new or virgin wool fibers; (3) percentages of reprocessed or reused fibers; (4) percentage of each fiber other than wool, if such fibers constitute 5% or more of the total; (5) aggregate of other fibers, (6) nonfibrous loading, filling or adulterating substance

The term wool, according to government standards, must always mean new wool that has not been made up in any form of wool product. New wool comes directly from a fleece. It has never been previously spun, woven, felted, or worn.

Reprocessed wool. According to the government classification, reprocessed wool is that which has been reclaimed and remanufactured from unused wool materials Such materials may be combings and scraps of wool obtained during the manufacturing processes, sample swatches, or pieces of all wool cloth from apparel manufacturing.

Reused wool. The government gives the special classification of reused wool to fiber salvaged from all kinds of used consumers' goods.

Virgin wool This term is now used by the textile industry to designate new wool from a sheep's fleece, but the term is too all-inclusive to serve as a criterion of quality. Although the term testifies to the fact that virgin wool does not contain remanufactured wool fibers, it does not distinguish between the less desirable fibers of a fleece and a specially fine quality of wool. Virgin wool may also include pulled or dead wool, which may be of definitely inferior quality. One should not feel that a fabric labeled "100% new wool" is necessarily more serviceable than that containing any of the remanufactured wool fibers.

Wool of different grades may be blended or mixed together. It is not uncommon for inferior grades to be mixed with the better grades The use of a mixture with a coarser grade of fiber is a legitimate practice if the purpose is to make a better wearing and less expensive product, provided the label on the finished goods indicates a true description of the raw materials used *See* ALPACA; CAMEL'S HAIR, CASHMERE; GOAT; LLAMA, MOHAIR; SHEEP, VICUNA [M D P]

Bibliography· *See* AGRICULTURAL SCIENCE (ANIMAL).

Work

In physics, the term work refers to the transference of energy that occurs when a force is applied to a body that is moving in such a way that the force has a component in the direction of the body's motion Thus work is done on a weight that is being lifted, or on a spring that is being stretched or compressed, or on a gas that is undergoing compression in a cylinder.

When the force acting on a moving body is constant in magnitude and direction, *the amount of work done* is defined as the product of just two factors: (1) the component of the force in the direction of motion, and (2) the distance moved by the point of application of the force. Thus the defining equation for work W is

$$W = f \cos \phi \cdot s \qquad (1)$$

where f and s are the magnitudes of the force and displacement, respectively, and ϕ is the angle between these two vector quantities (Fig. 1). Because $f \cos \phi \cdot s = f \cdot s \cos \phi$, work may be de-

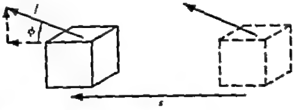


Fig 1 The work of a constant force f is $fs \cos \phi$

fined alternatively as the product of the force and the component of the displacement in the direction of the force In Fig. 2 the work of the constant force f when the application point moves along the curved path from P to P', and therefore undergoes the displacement $\overline{PP'}$, is $f \cdot \overline{PP'} \cos \phi$, or $f \cdot \overline{PE}$.

Work is a scalar quantity. Consequently, to find the total work done on a moving body by several different forces, the work of each may be computed separately and the ordinary algebraic sum taken.

[illegible]

a constant magnitude of 50 pounds of force (50 lbf) and let Eq (1) be used to compute the work W done under each of the following circumstances. (1) if the man pushes straight forward, in the direction of the car's displacement, then $\phi = 0°$, $\cos \phi = 1$, and $W = 50 \text{ lbf} \times 1 \times 30 \text{ ft} = 1500$ foot-pounds of force (ft-lbf); (2) if he pushes

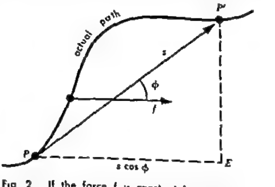


Fig 2 If the force f is constant in *magnitude* and direction, the work done in traversing any path connecting P and P' is $f \cdot \overline{PE}$.

Fig. 8. Wool coring

Action of cold At subzero temperatures, wool remains pliable and undergoes no perceptible chemical change.

Action of water and steam. Normally, wool is insoluble in water, though highly hygroscopic If wool is boiled for 2 hours, a weight loss of about 25% occurs. Conversely, the fiber diameter will swell approximately 10% with no damaging effects during brief periods of steaming (212°F). Wool is more efficiently converted into yarn when it has been kept in a warm, moist environment; however, prolonged steaming causes loss of strength. Wool is not allowed to become absolutely dry. Usually 12–16% of the moisture is left in the wool to condition it for subsequent handling

Plasticity. Under moist conditions wool becomes plastic. Its shape can be altered and its affinity for dye changed. Knowledge of this characteristic enables the manufacturer to set yarns and to produce desired color effects Plasticity increases with rising temperatures to a point at which stretched fibers become permanently set and will not return to their normal state.

Table 5. Amino acid composition of wool*

Amino acid	Approximate percentage present in wool	Grams of residue per 100 g of wool	Grams of side-chain per 100 g of wool
Glycine	6 5	4 94	0 09
Alanine	4 4	3 52	0 74
Serine	9.41	7 80	2 76
Proline	6 75	5 69	2.46
Valine	4 72	3 99	1.73
Threonine	6 76	5 74	2 59
Cystine	12 72†	10 83	4 89
Leucine isomers	11 3	9.75	4 92
Aspartic acid	7 27	6 28	3 22
Lysine	3 3	2 89	1 63
Glutamic acid	15 27	13 40	7 58
Methionine	0 71	0 62	0 36
Histidine	0 7	0 62	0 37
Hydroxylysine	0 21	0 19	0 11
Phenylalanine	3 75	3 34	2 07
Arginine	10 4	9 33	5 97
Tyrosine	5 8	5 23	3 43
Tryptophan	0 7	0 64	0 45
Total	110 67	94 80	45 37
Ammonia nitrogen	1.18	−0 30	−0 30
Total, corrected for ammonia nitrogen		94 50	45 07

* From J M Matthews

† Based on 3 55% total sulfur and subtracting methionine sulfur

Technological advances. The inherent advantages of wool are being exploited, and its limitations as a textile fiber are being overcome by the application of technology to manufacturing processes The use of the insecticide dieldrin as a dye renders wool mothproof for life Permanent pleats have been imparted to garments which are shrinkproofed and can be home laundered. Each such technological advance enables wool to hold its competitive place in the field of textile manufacture and use. [T.D.W.]

Reuse of wool fibers. There has never been a sufficient supply of new wool stocks to take care of a steadily increasing demand for wool To meet this situation, wool fibers have had to be recovered from old clothing, rags of all kinds, and waste from wool manufacturing This wool is variously called salvaged, reclaimed, reworked, or remanufactured, but it is best known in the textile industry as shoddy. This term is misunderstood by the average consumer, who is inclined to believe that wool fabric containing remanufactured fibers is necessarily of inferior quality.

The hardier, although less resilient, remanufactured fibers, when obtained from good original stock and combined with new wool from lambs, add durability to the soft new wool. Thus, remanufactured fibers contribute ability to withstand hard wear, although there is some sacrifice in warmth, softness of texture, and resiliency. They also make wool clothing available at lower prices.

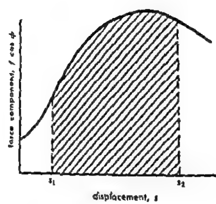


Fig 4. A work diagram.

sented by the area under the resulting curve between s_1 and s_2 and can be computed by measuring this area, due allowance being made for the scale in which the diagram is drawn

For an infinitely small displacement ds of the point of application of the force, the increment of work dW is given by

$$dW = f \cos \phi \, ds \tag{2}$$

a differential expression that provides the most general definition of the concept of work. In the language of vector analysis, dW is the scalar product of the vector quantities **f** and d**s**, and Eq (2) then takes the form $dW = \mathbf{f} \cdot d\mathbf{s}$ If the force is a known continuous function of the displacement, the total work done in a finite displacement from point P to point P' of the path is obtained by evaluating the line integral

$$W = \int_P^{P'} f \cos \phi \, ds = \int_P^{P'} \mathbf{f} \cdot d\mathbf{s} \tag{3}$$

When a variable torque of magnitude τ acts on a body mounted on a fixed axis, the work done is given by $W = \int_{\theta_1}^{\theta_2} \tau \, d\theta$, where $\theta_2 - \theta_1$ is the total angular displacement expressed in radians. See ENERGY

[D E R]

Bibliography. R A. Becker, *Introduction to Theoretical Mechanics*, 1954, G Holton and D H D Roller, *Foundations of Modern Physical Science*, 1958, M W Zemansky *Heat and Thermodynamics*, 4th ed , 1957

Work function (electronic)

A quantity with the dimensions of energy which determines the thermionic emission of a solid at a given temperature The thermionic electron current density J emitted by the surface of a hot conductor at a temperature T is given by the Richardson formula

$$J = AT^2 e^{-\phi/kT}$$

where A is a constant, k is Boltzmann's constant ($\approx 1.38 \times 10^{-23}$ joule per degree centigrade) and ϕ is the work function, the latter may be determined from a plot of log (J/T^2) versus $1/T$ (see THERMIONIC EMISSION) For metals, ϕ may also be determined by measuring the photoemission as a function of the frequency of the incident electromagnetic radiation, ϕ is then equal to the minimum (threshold) frequency for which electron emission is observed times Planck's constant h ($\approx 6.65 \times 10^{-34}$ joule sec). The work function of a solid is usually expressed in electron volts (1 ev is the energy gained by an electron as it passes through a potential difference of 1 volt, and is equal to 1.60×10^{-19} joule). A list of average values of work functions (in electron volts) for metals follows.

Al	4.20	Cs	1.93	Na	2.28
Ag	4.46	Cu	4.45	Ni	4.96
Au	4.89	Fe	4.44	Pd	4.98
Ba	2.51	K	2.22	Pt	5.36
Cd	4.10	Li	2.48	Ta	4.13
Co	4.41	Mg	3.67	W	4.54
Cr	4.60	Mo	4.24	Zn	4.29

The work function of metals varies from one crystal plane to another and also varies slightly with temperature (approximately 10^{-4} ev/degree). For a metal, the work function has a simple interpretation. At absolute zero, the energy of the most energetic electrons in a metal is referred to as the Fermi energy, the work function of a metal is then equal to the energy required to raise an electron with the Fermi energy to the energy level corresponding to an electron at rest in vacuum. The work function of a semiconductor or an insulator has the same interpretation, but in these materials the Fermi level is in general not occupied by electrons and thus has a more abstract meaning. See FIELD EMISSION, PHOTOEMISSION. [A.J DE.]

Bibliography: See THERMIONIC EMISSION.

Work function (thermodynamics)

A thermodynamic function, also called the work content, Helmholtz free energy, or by the European school, simply the free energy. It is defined as the internal energy E of a system minus the temperature-entropy product, TS, and has a characteristic value for each state of a system. In an isothermal process, the maximum work which can be done by a system is equal to the decrease in its work function When only work due to expansion against a fluid pressure is possible, as in ordinary chemical reactions, a spontaneous process at *constant temperature* and volume is characterized by a decrease in the work function, whereas the corresponding *criterion for equilibrium* is that the work function for the system should be at a *minimum*. See THERMODYNAMICS (CHEMICAL).

[P.B.]

Work measurement

Scientific management is based on measurement plus control Without measurement there can be little control. There must be a standard before anyone—management, supervisor, or worker—can judge how well . . .

in a sidewise direction making an angle ϕ of 60° with the displacement, then $\cos 60° = 0.50$ and $W = 750$ ft-lbf; (3) if he pushes against the side of the car and therefore at right angles to the displacement, $\phi = 90°$, $\cos \phi = 0$, and $W = 0$; (4) if he pushes or pulls backward, in the direction opposite to the car's displacement, $\phi = 180°$, $\cos \phi = -1$, and $W = -1500$ ft-lbf.

Notice that the work done is positive in sign whenever the force or any component of it is in the same direction as the displacement; one then says that work is being done *by* the agent exerting the force (in the example, the man) and *on* the moving body (the car). The work is seen to be negative whenever the direction of the force or force component is opposite to that of the displacement; then work is said to be done *on* the agent (the man) and *by* the moving body (the car). From the point of view of energy, an agent doing positive work is losing energy to the body on which the work is done, and one doing negative work is gaining energy from that body.

Units of work and energy. These consist of the product of any force unit and any distance unit. Units in common use are the foot pound, the foot poundal, the erg, and the joule. The product of any power unit and any time unit is also a unit of work or energy. Thus the horsepower hour (hp-hr) is equivalent, in view of the definition of the horsepower, to 550 ft-lbf/sec × 3600 sec, or 1,980,000 ft-lbf. Similarly, the watt-hour is 1 joule/sec × 3600 sec, or 3600 joule; and the kilowatt-hour is 3,600,000 joule. *See* ERG; FOOT-POUND; FOOT-POUNDAL; HORSEPOWER; JOULE; WATT

Work of a torque. When a body which is mounted on a fixed axis is acted upon by a constant torque of magnitude τ and turns through an angle θ (radians), the work done by the torque is $\tau\theta$.

Work principle. This principle, which is a generalization from experiments on many types of machines, asserts that, during any given time, the work of the forces applied to the machine is equal to the work of the forces resisting the motion of the machine, whether these resisting forces arise from gravity, friction, molecular interactions, or inertia. When the resisting force is gravity, the work of this force is mgh, where mg is the weight of the body and h is the vertical distance through which the body's center of gravity is raised. Note that if a body is moving in a horizontal direction, h is zero and no work is done by or against the gravitational force of the earth. If a person holds an object or carries it across level ground, he does no *net* work against gravity; yet he becomes fatigued because his tensed muscles continually contract and relax in minute motions, and in walking he alternately raises and lowers the object and himself.

The resisting force may be due to molecular forces, as when a coiled elastic spring is being compressed or stretched. From Hooke's law, the average resisting force in the spring is $-\frac{1}{2}ks$, where k is the force constant of the spring and s is the displacement of the end of the spring from its normal position; hence the work of this elastic force is $-\frac{1}{2}ks^2$. *See* HOOKE'S LAW.

If a machine has any part of mass m that is undergoing an acceleration of magnitude a, the resisting force $-ma$ which the part offers because of its inertia involves work that must be taken into account; and similarly for the resisting torque $-I\alpha$ if any rotating part of moment of inertia I undergoes an angular acceleration α.

When the resisting force arises from friction between solid surfaces, the work of the frictional force is $-\mu f_n s$, where μ is the coefficient of friction for the pair of surfaces, f_n is the normal force pressing the two surfaces together, and s is the displacement of the one surface relative to the other during the time under consideration. The frictional force μf_n and the displacement s giving rise to it are always opposite in direction ($\phi = 180°$). *See* FRICTION

The work done by any conservative force, such as a gravitational, elastic, or electrostatic force, during a displacement of a body from one point to another has the important property of being path-independent: its value depends only on the initial and final positions of the body, not upon the path traversed between these two positions. On the other hand, the work done by any nonconservative force, such as friction due to air, depends on the path followed and not alone on the initial and final positions, for the direction of such a force varies with the path, being at every point of the path tangential to it. *See* FORCE.

Work of a variable force. If the force varies in magnitude and direction along the path $\overline{PP'}$ of its point of application, one must first divide the whole path into parts of length Δs, each so short that the force component $f \cos \phi$ may be regarded as constant while the point of application traverses it (Fig. 3). Equation (1) can then be applied to each small part and the resulting increments of work added to get the total work done. Various devices are available for measuring the force component as a function of position along the path. Then a *work diagram* can be plotted (Fig. 4). The total work done between positions s_1 and s_2 is repre-

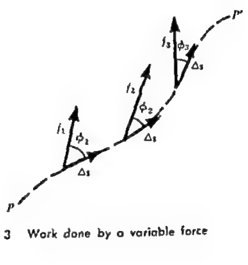


Fig. 3 Work done by a variable force

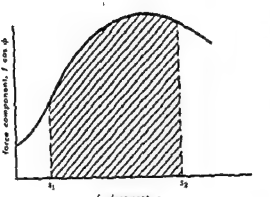


fig 4 A work diagram

sented by the area under the resulting curve between s_1 and s_2 and can be computed by measuring this area, due allowance being made for the scale in which the diagram is drawn

For an infinitely small displacement ds of the point of application of the force the increment of work dW is given by

$$dW = f \cos \phi \; ds \qquad (2)$$

a differential expression that provides the most general definition of the concept of work In the language of vector analysis, dW is the scalar product of the vector quantities $\mathbf{f}$ and $d\mathbf{s}$, and Eq (2) then takes the form $dW = \mathbf{f} \cdot d\mathbf{s}$. If the force is a known continuous function of the displacement, the total work done in a finite displacement from point P to point P' of the path is obtained by evaluating the line integral

$$W = \int_P^{P'} f \cos \phi \; ds = \int_P^{P'} \mathbf{f} \cdot d\mathbf{s} \qquad (3)$$

When a variable torque of magnitude τ acts on a body mounted on a fixed axis, the work done is given by $W = \int_{\theta_1}^{\theta_2} \tau \, d\theta$, where $\theta_2 - \theta_1$ is the total angular displacement expressed in radians See ENERGY

[D.E R]

Bibliography R A Becker, *Introduction to Theoretical Mechanics*, 1954. G Holton and D H D Roller, *Foundations of Modern Physical Science*, 1958 M W Zemansky, *Heat and Thermodynamics*, 4th ed , 1957

Work function (electronic)

A quantity with the dimensions of energy which determines the thermionic emission of a solid at a given temperature The thermionic electron current density J emitted by the surface of a hot conductor at a temperature T is given by the Richardson formula

$$J = AT^2 e^{-\phi/kT}$$

where A is a constant, k is Boltzmann's constant ($= 1.38 \times 10^{-23}$ joule per degree centigrade) and ϕ is the work function, the latter may be determined from a plot of log (J/T^2) *versus* $1/T$ (see THERMIONIC EMISSION) For metals, ϕ may also be determined by measuring the photoemission as a function of the frequency of the incident electromagnetic radiation; ϕ is then equal to the minimum (threshold) frequency for which electron emission is observed times Planck's constant h ($= 6.65 \times 10^{-34}$ joule sec). The work function of a solid is usually expressed in electron volts (1 ev is the energy gained by an electron as it passes through a potential difference of 1 volt, and is equal to 1.60×10^{-19} joule). A list of average values of work functions (in electron volts) for metals follows.

Al	4 20	Cs	1.93	Na	2.28
Ag	4 46	Cu	4.45	Ni	4 96
Au	4 89	Fe	4 44	Pd	4.98
Ba	2.51	K	2.22	Pt	5 36
Cd	4 10	Li	2.48	Ta	4.13
Co	4.41	Mg	3 67	W	4 54
Cr	4 60	Mo	4 24	Zn	4.29

The work function of metals varies from one crystal plane to another and also varies slightly with temperature (approximately 10^{-4} ev/degree). For a metal, the work function has a simple interpretation At absolute zero, the energy of the most energetic electrons in a metal is referred to as the Fermi energy; the work function of a metal is then equal to the energy required to raise an electron with the Fermi energy to the energy level corresponding to an electron at rest in vacuum. The work function of a semiconductor or an insulator has the same interpretation, but in these materials the Fermi level is in general not occupied by electrons and thus has a more abstract meaning. See FIELD EMISSION, PHOTOEMISSION. [A J DE.]

Bibliography: See THERMIONIC EMISSION

Work function (thermodynamics)

A thermodynamic function, also called the work content, Helmholtz free energy, or by the European school, simply the free energy It is defined as the internal energy E of a system minus the temperature-entropy product, TS, and has a characteristic value for each state of a system. In an isothermal process the maximum work which can be done by a system is equal to the decrease in its work function When only work due to expansion against a fluid pressure is possible, as in ordinary chemical reactions, a spontaneous process at constant temperature and volume is characterized by a decrease in the work function, whereas the corresponding criterion for equilibrium is that the work function for the system should be at a *minimum* See THERMODYNAMICS (CHEMICAL) [P B.]

Work measurement

Scientific management is based on measurement plus control Without measurement there can be little control. There must be a standard before any-one—management, supervisor, or worker—can judge how well a job is done.

Informal standards based on estimate, records of past performance, or mere[illegible]

are inconsistent and inaccurate. They cause many industrial relations problems. To avoid these problems, standards based on work measurement are used.

Devices *and procedures for work measurement* include stop watch, motion picture camera, microchronometer, specially designed electronic computers, predetermined elemental time systems, and work sampling. Of these, stop watch time study, predetermined elemental time systems, and work sampling are the most widely used.

Time study. *Actual elapsed time for performing* an operation or subdivision thereof is determined by a suitable timing device and recorded. The procedure often includes adjustment of the actual time as the result of performance rating to derive the time which should be required to perform the task by a workman working at a standard pace and following a standard method under standard conditions. Time studies are made in industry by men variously called industrial engineers, methods engineers, time-study engineers, or time-study men. The last term is perhaps the most common.

If the worker who is being studied uses the correct work method, accurate measurement with a stop watch of the required time to do the job is relatively simple. If the worker uses an incorrect method, it is almost impossible to establish accurately the time for doing the same task by a different method. As obvious as this fact is, the failure to recognize it or to act upon it is responsible for many of the inaccurate standards that exist in industry.

Figure 1 is a graphic analysis of the elements of time study. At the outset the time-study man selects a qualified operator to study. He explains the purpose of the study and tries to establish a cooperative relationship with the workman. He then observes the operation as the worker performs it. He subdivides the total work cycle into smaller *elements and records descriptions of the elements* on a prepared time-study form. Then, standing where he can clearly observe the operator without getting in his way, he starts his stop watch at the beginning of the first element to be included in the study and records the time of day. As each element ends and the next one begins, he records the time. *If the worker introduces a foreign element or performs* elements out of sequence, the time-study man notes this fact.

The time-study man allows his watch to run continuously and observes sufficient pieces to insure a representative sample. At the end of the study, he again notes the time of day. He judges the skill *and effort exhibited by the worker while he* was being timed and records his conclusions. After recording all information needed to identify the study, including the operator, part number, ma-

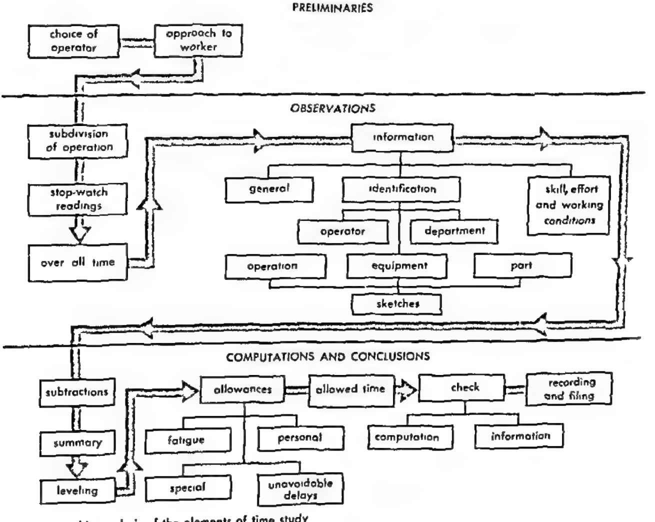


Fig. 1. Graphic analysis of the elements of time study

time number, and the like, he computes the time taken to perform each element. He discards times that seem abnormal and averages the remaining elapsed times for each element. He then adjusts them by a procedure known as leveling, in accordance with the rated performance of the operator. If the worker worked faster than normal, he increases the averaged elapsed times correspondingly and vice versa. Table 1 shows a widely used

Table 1 Performance rating factors

Skill		Effort	
A_1 = +0.15	Super skill	A_1 = +0.13	Excessive effort
A = +0.14		A = +0.125	
A_2 = +0.13		A_2 = +0.12	
B_1 = +0.11	Excellent skill	B_1 = +0.10	Excellent effort
B = +0.095		B = +0.09	
B_2 = +0.08		B_2 = +0.08	
C_1 = +0.06	Good skill	C_1 = +0.05	Good effort
C = +0.045		C = +0.035	
C_2 = +0.03		C_2 = +0.02	
D = 0.00	Average	D = 0.00	Average
E_1 = −0.05	Fair skill	E_1 = −0.04	Fair effort
E = −0.075		E = −0.06	
E_2 = −0.10		E_2 = −0.08	
F_1 = −0.16	Poor skill	F_1 = −0.12	Poor effort
F = −0.19		F = −0.145	
F_2 = −0.22		F_2 = −0.17	

table of performance-rating factors for skill and effort. These values added algebraically to 1.0 give the leveling factor which is used to adjust the raw time data to the average performance level.

The time-study man next adds allowances for fatigue and personal, unavoidable, and special delays. He then adds the elemental standard times multiplied by the total number of occurrences to determine the total time standard for the operation. After checking his work for accuracy, he records the established standard.

Predetermined elemental time. The most widely known and used predetermined elemental time systems are Methods-Time Measurement or MTM, Work Factor system or WF, Basic Motion Time study or BMT, and Motion-Time-Analysis or MTA. Of these, MTM was the first to be published; it will be used to illustrate work measurement by predetermined elemental times.

MTM analyzes any manual operation or method into the basic motions required to perform it and assigns to ea[illegible]

ard. Th[illegible]

of the m[illegible]

made [illegible] measurement is only one of its many uses. Others are (1) developing effective methods in advance of production, (2) improving existing methods, (3) establishing time standards, (4) developing time formulas and standard data, (5) estimating, (6) guiding product design, (7) developing effective tool designs, (8) selecting effective equipment, (9) training supervisors to become methods conscious, (10) settling grievances, (11) operator training, and (12) research, particularly in connection with methods, learning time, and performance rating.

From analysis of motion pictures of many qualified operators, the times required to perform various industrial motions were determined. Table 2

Table 2. MTM data for position motions of Fig. 2*

Class of fit	Symmetry	Easy to handle	Difficult to handle
Loose (No pressure required)	S	5.6	11.2
	SS	9.1	14.7
	NS	10.4	16.0
Close (Light pressure required)	S	16.2	21.8
	SS	19.7	25.3
	NS	21.0	26.6
Exact (Heavy pressure required)	S	43.0	48.6
	SS	46.5	52.1
	NS	47.8	53.4

* Distance moved to engage 1 in. or less

shows such times to position two hand-held objects in relation to each other when the objects are moved into engagement an inch or less (Fig. 2). As in the other motions, several major variables affect time to position objects: class of fit, symmetry, and ease of handling. Fit and symmetry are de-

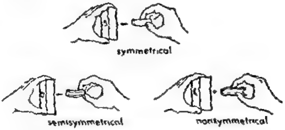


Fig. 2 MTM times to position an object under three conditions of symmetry are given in Table 2

fined in the table and figure; ease of handling is influenced by the manner in which the object is grasped and the nature of the object itself. An object is easy to handle if it is grasped near the point of engagement (and need not be regrasped during positioning) and if the object is dimensionally stable. Small parts or long flexible parts are difficult to handle. The tabulated times are in hundred thousandths of an hour; that is, 1 TMU = 0.00001 hr.

To establish a time standard using tabulated MTM data, the methods engineer breaks the activities into standard motions. This he does by observing an operator who is doing the job or by visualizing the motions which should be used. If the method he records is ineffective, he corrects the motion pattern before proceeding.

When an acceptable method has been recorded, he determines the time required for each motion from previously established standards. They represent the time required by an operator of average skill working with average effort to make the motions, thus performance rating is eliminated. The

time for performing the operation. When this is increased by allowances for fatigue and for personal, unavoidable, and special delays, the result is the standard time for the job.

On repetitive work, the most important use of MTM is for methods improvement, with work measurement being secondary because it can be accomplished by other methods such as time study for about the same cost. On nonrepetitive work, the situation is reversed. MTM makes possible the development of time formulas or standard data in one *quarter of the time required by time study. This is* because it is unnecessary to observe and time every condition which must be considered when compiling a formula. The time can be obtained by visualizing the method which will be used to cope with each situation. Thus MTM makes it economical to apply standards to nonrepetitive work and to indirect labor. *In this case, its methods-study feature* is of secondary importance because it is impractical to attempt to develop and train workers to use highly refined methods on nonrepetitive jobs.

MTM, in common with some of the other predetermined elemental time systems, provides a simplified set of data to use where quickness of appli*cation is preferable to complete accuracy. The sim*plified data are shown by Table 3 They include a 15% allowance for fatigue and personal and unavoidable delays.

Table 3. Simplified methods-time measurement application data*

		TMU		TMU
Reach or move			Body, leg, and eye motions	
Type 1		3 + length	Simple foot motion	10
Types 2 and 3		length	*Foot motion with pressure*	20
Weight		1 per 2 lb	Leg motion	10
Position	S	NS SS	Side step case 1	20
Loose	10	15	Side step case 2	40
Close	20	25	Turn body case 1	20
Exact	50	55	Turn body case 2	45
Grasp			Eye time	10
Simple		3	Bend	35
Regrasp or transfer		6	*Stoop*	35
Complex		10	Kneel on one knee	35
Disengage			Arise	35
Loose		5	Kneel on both knees	80
Close		10	Arise	90
Exact		30	Sit	40
Release		2	Stand	50
Turn		8	Walk per pace	17
Apply pressure		15		

* All times include 15% allowance. 1 TMU = .00001 hr = .0006 min = .036 sec

Work sampling. This statistical sampling technique is used to determine the proportions of delays and other classifications of activity present in the total work day. Work sampling was originally called ratio delay because it was used to determine the amount of and causes for downtime on textile machines

The assumption behind work sampling is that the number of occurrences of an event, expressed in per cent, as determined by a number of random observations is the same as the time taken by the event, expressed in per cent This relationship holds true regardless of the length or frequency of the event.

Work sampling gives useful information about the per cent of the day that is used for each activity that makes up a job or class of work. Delay times can be determined, as well as the time for any other event, *whether it be time taken for setup,* man time as compared to machine time, or the time for any element of productive work. The technique may thus be used as a basis for job analysis or to determine departmental efficiencies, multi-crew work loads, machine or man utilization, or in any other situation where a breakdown of the work into *the time consumed* by each classification of activity will be useful.

The most commonly used alternative to collecting information by work sampling is the all-day time study. The methods engineer *observes a* worker for a full day or for several days, recording everything he does and the time he takes to do it. The advantages of the work sampling method over the all-day time study are (1) the information is obtained at lower cost, (2) the procedure is more acceptable to workers and supervisors, (3) the procedure is more convenient for the methods engineer, (4) as many observers as desired can make the study, (5) no stop watch is needed, (6) more representative data are obtained, and (7) many studies may be made simultaneously.

Work sampling is based on the laws of proba*bility. In work-sampling studies, there are four* quantities that must be considered. The first is the per cent occurrence of an element. This is the quantity sought The second is the error which can be tolerated This can be found from $x = yp$ where x is acceptable tolerance, y is accuracy desired, and p is per cent occurrence of the element The third is the per cent of time the answer must be within the acceptable tolerance, computed from the formula

$$S = \sqrt{\frac{p(1-p)}{N}}$$

where S is the standard error expressed as a decimal and N is total number of observations, the fourth quantity that must be considered In other words, the methods engineer must know how many observations to make to obtain the accuracy he desires Figure 3 shows in graphic form the elements of a work-sampling study

Time formulas and standard data. A time formula is an algebraic expression of the factors de*termining the time required to perform* an operation. It is a set of element time data for a given class of work reduced to its simplest form This form may be a single table or curve, a combination of tables and curves, or *the more conventional* algebraic equation. Data compiled as tables, charts,

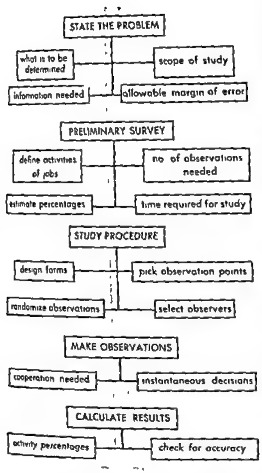


Fig 3 Graphic analysis of the elements of work sampling

or lists of element times are generally known as standard data

The objective of time formulas is to enable a trained person to determine the standard time for performing an operation within a given class of work if he knows the *characteristics* of the part on which the work is done The methods engineer determines which elements are always performed, which are performed only on some occasions, which are *constant*, and which are variable and why *This analysis* is made once, when the formula is derived Thus the *determination* of a standard from a formula is a quick, *routine* task Standards are consistent for a *given* class of work, because they are all determined *from the same* formula The *computation* of standards may be done by a well-trained clerk, thus relieving the methods engineer of routine work Because of the low cost of establishing standards by this method, time *formulas* make feasible the measurement and *control* of low quantity productive work and indirect labor operations such as receiving and shipping storekeeping tool making, and maintenance

Time data covering representative jobs within a given class of work are obtained either from time studies or, more quickly and in more convenient form, from studies made by a predetermined elemental time system. These data are posted on a columnar sheet known as the master table of detailed time studies or more simply as a spread sheet. The data are analyzed, constants and variables are identified, and the occasion and frequency of the use of each element are determined. The data are then compiled into the simplest form for subsequent application. A formula report is written describing (1) parts covered by formula, (2) the operation, (3) the work station, (4) the formula itself, (5) its area of application, (6) working methods used and working conditions, (7) the data used in deriving the formula, (8) the details of the derivation, (9) quality requirements of the work, and if appropriate (10) the wage payment plan under which the standards are applicable. Finally, the papers accumulated during the process are identified and filed so that they will be available for reference if questions about the formula are *raised in the future.*

To establish a time standard from a time formula, the time-study man determines the variable characteristics of the part to be worked on from a drawing or the part itself He then substitutes these *in the formula and computes the standard It* requires usually from 1–15 min to establish a standard by this method, in most cases it takes less than 5 min *See* METHODS ENGINEERING [H B.M.]

Bibliography: S. M Lowry, H B Maynard, and G J Stegemerten, *Time and Motion Study*, 3d ed, 1940, H B. Maynard, G J. Stegemerten, and J. L. Schwab, *Methods-Time Measurement*, 1948.

Work standardization

The general concept of standardization applied to the performance of jobs or operations in industry or business. It establishes uniformity of working conditions tools, equipment, technical procedures, administrative procedures, workplace arrangements, motion sequences, materials, quality requirements, and similar factors which affect the performance of work (*see* DESIGN STANDARDS).

Work standardization is one of the specific steps performed in the field of methods engineering which is, in turn, part of the field of industrial engineering Objectives of work standardization are lower costs, greater productivity, *improved* quality of workmanship, greater safety, and quicker and better development of skills among workers Effort to standardize work is usually made by industrial engineers, but it may be made by anyone connected with the work.

There is no specific technique or procedure for putting the concept of work standardization into effect. Engineers depend largely on their awareness of the concept, their energy and ingenuity in applying it, and their breadth of experience with the particular jobs or operations in their own businesses.

Any planned effort to standardize a job, however, must start with a systematic way of describing or cataloging jobs or operations in terms of their essential characteristics (*see* METHODS

time for performing the operation. When this is increased by allowances for fatigue and for personal, unavoidable, and special delays, the result is the standard time for the job.

On repetitive work, the most important use of MTM is for methods improvement, with work measurement being secondary because it can be accomplished by other methods such as time study for about the same cost. On nonrepetitive work, the situation is reversed. MTM makes possible the development of time formulas or standard data in one quarter of the time required by time study. This is because it is unnecessary to observe and time every condition which must be considered when compiling a formula. The time can be obtained by visualizing the method which will be used to cope with each situation. Thus MTM makes it economical to apply standards to nonrepetitive work and to indirect labor. In this case, its methods-study feature is of secondary importance because it is impractical to attempt to develop and train workers to use highly refined methods on nonrepetitive jobs.

MTM, in common with some of the other predetermined elemental time systems, provides a simplified set of data to use where quickness of application is preferable to complete accuracy The simplified data are shown by Table 3 They include a 15% allowance for fatigue and personal and unavoidable delays.

Table 3. Simplified methods-time measurement application data*

		TMU		TMU
Reach or move			Body, leg, and eye motions	
Type 1		3 + length		
Types 2 and 3		length	Simple foot motion	10
Weight		1 per 2 lb		
Position			Foot motion with pressure	20
	S	NS SS	Leg motion	10
Loose	10	15	Side step case 1	20
Close	20	25	Side step case 2	40
Exact	50	55	Turn body case 1	20
Grasp			Turn body case 2	45
Simple		3	Eye time	10
Regrasp or transfer		6	Bend	35
Complex		10	Stoop	35
Disengage			Kneel on one knee	35
Loose		5	Arise	35
Close		10	Kneel on both knees	80
Exact		30	Arise	90
Release		2	Sit	40
Turn		8	Stand	50
Apply pressure		15	Walk per pace	17

* All times include 15% allowance 1 TMU = 00001 hr = 0006 min = 036 sec

Work sampling. This statistical sampling technique is used to determine the proportions of delays and other classifications of activity present in the total work day. Work sampling was originally called ratio delay because it was used to determine the amount of and causes for downtime on textile machines.

The assumption behind work sampling is that the number of occurrences of an event, expressed in per cent, as determined by a number of random observations is the same as the time taken by the event, expressed in per cent This relationship holds true regardless of the length or frequency of the event.

Work sampling gives useful information about the per cent of the day that is used for each activity that makes up a job or class of work. Delay times can be determined, as well as the time for any other event, whether it be time taken for setup, man time as compared to machine time, or the time for any element of productive work. The technique may thus be used as a basis for job analysis or to determine departmental efficiencies, multi-crew work loads, machine or man utilization, or in any other situation where a breakdown of the work into the time consumed by each classification of activity will be useful.

The most commonly used alternative to collecting information by work sampling is the all-day time study. The methods engineer observes a worker for a full day or for several days, recording everything he does and the time he takes to do it. The advantages of the work sampling method over the all-day time study are (1) the information is obtained at lower cost, (2) the procedure is more acceptable to workers and supervisors, (3) the procedure is more convenient for the methods engineer, (4) as many observers as desired can make the study, (5) no stop watch is needed, (6) more representative data are obtained, and (7) many studies may be made simultaneously.

Work sampling is based on the laws of probability. In work-sampling studies, there are four quantities that must be considered The first is the per cent occurrence of an element. This is the quantity sought The second is the error which can be tolerated. This can be found from $x = yp$ where x is acceptable tolerance, y is accuracy desired, and p is per cent occurrence of the element The third is the per cent of time the answer must be within the acceptable tolerance, computed from the formula

$$S = \sqrt{\frac{p(1-p)}{N}}$$

where S is the standard error expressed as a decimal and N is total number of observations, the fourth quantity that must be considered In other words, the methods engineer must know how many observations to make to obtain the accuracy he desires. Figure 3 shows in graphic form the elements of a work-sampling study

Time formulas and standard data. A time formula is an algebraic expression of the factors determining the time required to perform an operation. It is a set of element time data for a given class of work reduced to its simplest form This form may be a single table or curve, a combination of tables and curves, or the more conventional algebraic equation. Data compiled as tables, charts,

aedon, is a familiar yard and garden bird, readily accepting nesting boxes and unafraid of man. It nests over the northern two-thirds of the United States and across southern Canada. The winter wren, *T. troglodytes*, is Holarctic in distribution, nesting in the boreal forests of Eurasia and America, and in the United States wintering southward to the Gulf Coast. Few birds are as closely restricted to a limited habitat as this minute, short-tailed wren is to the brush pile, fallen tree environment. The large, long-tailed cactus wren, *Campylorhynchus brunneicapillus*, is characteristic of the Southwestern desert. Several other species are relatively common in the United States. *See* PASSERIFORMES. [J.D.B.]

Wrist

The flexible link between *forearm and hand* consisting of the wrist joint and adjacent structures. This articulation site is comprised of a set of synovial joints formed by the *head of the radius* and four upper carpals, or wrist bones. The four lower carpals articulate with the upper hand bones, the metacarpals. All bony structures *are held in place* by deep ligaments, but movement is obtained by forearm muscles acting through tendons passing to the hand and fingers. Other subcutaneous structures include the radial and ulnar arteries, superficial and deep veins, and branches of the radial, median, and ulnar nerves.

Human wrist, volar aspect. (From C. M. Goss, ed., *Gray's Anatomy of the Human Body, 22nd ed.*, Lea and Febiger, 1930)

Analogous structures appear in all tetrapods but they are more complex in those which have developed the paw or hand for special activities. *See* SKELETAL SYSTEM. [E.C.ST.]

Wrought iron

As defined by the American Society for Testing Materials, wrought iron is "A ferrous material, aggregated from a solidifying mass of pasty particles of highly refined metallic iron, with which, without subsequent fusion, is incorporated a minutely and uniformly distributed quantity of slag."

This slag is a ferrous silicate resulting from the oxidizing reactions of refining, and it varies in amount from 1–3% in various types of final product. It is in purely mechanical association with the iron base-metal, as contrasted with the alloying relationships of the metalloids present in steel.

Wrought iron had a dominant position in meeting the need for a forgeable, nonbrittle, ferrous metal, from before the time of historical record until the advent of the age of steel.

The earlier primitive methods were finally succeeded by the puddling furnace of H. Cort in 1784. Wrought iron thus produced played a major role in the industrial revolution until the steel-making inventions of H. Bessemer in 1856, and the open-hearth regenerative furnace of W. Siemens several years later.

Whether the product is wrought iron or wrought steel, similar chemical reactions are involved in refining a molten charge of pig iron to a composition approaching pure iron. The associated metalloids are oxidized—carbon to form a gas, and silicon, manganese, and phosphorus with appropriate additions to form a liquid, insoluble slag.

The vital demarcation between the puddling and steel-making furnaces was the limited temperature attainable in the former. It was below the 2700°F needed to keep the refined iron molten. The freezing point of the iron rises as refining proceeds. Practically all the 3–4% carbon in the original charge is removed, and the base metal progressively solidifies, or "comes to nature." Manual manipulation accompanied by internal reaction results in a spongy mass of virtually pure iron, impregnated with iron silicate slag. Subsequent squeezing and rolling operations eject most of the still-fluid slag, but the small percentage finally retained is elongated into the distributed threadlike structure that is characteristic of *quality* wrought iron.

In steel-making operations, on the contrary, the high temperature maintains fluidity of both metal and slag, with liquation and separation of the latter. The slag-free metal is poured into molds, ready for the subsequent rolling operations.

A [illegible]
soon [illegible]
as co [illegible]
allog[illegible] analysis in later years showed that this results from the fiberlike slag inclusions that are finely and uniformly intermingled with the iron base-metal.

The development of steel-making processes prompted prediction of the doom of wrought iron. Puddling required hard manual labor, with low output both in unit mass and total tonnage, and accompanying high cost. However, wrought iron held a recognized competitive merit in certain uses, notably where corrosion- and shock-resistance were important.

STUDY). For instance, a 5-digit number might be used to describe all operations. The first digit could describe the basic operation, such as painting, welding, or assembly. The second digit could describe the size of the parts involved, the third digit the shape of the parts, the fourth digit the kind of material involved, and the fifth digit the *type of tools required.*

With all work described or codified in this way, and with the use of modern punched-card or similar devices to analyze data, an engineer could identify all jobs or operations with similar characteristics and thus find places where the concept of work standardization might be used. He would then use his imagination and ingenuity to see what specific opportunities existed. Examples of such work standardization follow.

An office manager, studying the design of forms, finds that certain information is needed on several different forms. He redesigns the forms so that the information appears in the same place and in the same general way on each form. This change will reduce the number of clerical errors and the time required to teach procedures to new clerks.

An industrial engineer studies various jobs done by a group of machine repairmen in his maintenance department. He develops a standard kit of hand tools and provides each repairman with a kit so that all work will be done with the most effective tool. Delays due to failure to have proper tools are thereby reduced.

An industrial engineer has made a motion study of small assembly operations in his plant He finds that many small, flat parts must be picked up and positioned in various assemblies. He also finds that one type and sequence of motions permits this work to be done in the quickest and easiest way. *He then standardizes on this motion sequence for* doing this part of the work and teaches it to all operators.

Or the sequence of events in reaching a standard may be reversed as in the following examples.

On a given operation, an industrial engineer discovers that a new type of power-driven wrench saves time in tightening nuts He then provides this type of wrench for all other operations on which similar work is done.

A machinist discovers that a better finish is produced on certain metal when his cutting tool is sharpened in a particular way. His foreman then directs that all cutting tools used by all machinists on that metal be sharpened in like manner.

A safety director, after comparative tests, decides that a particular eye shield provides greater protection for workers against injury on certain types of work. He then sees that this shield is available and in use at all places in the plant where this type of work is done.

Work standardization differs from the establishment of time standards: work standardization concerns more how an operation is performed than with how long it takes. Usually, work standardization is completed before time standards are established. *See* METHODS ENGINEERING. [H.T.S.]

Worm

Any one of a large variety of slender, elongated, crawling or creeping invertebrate animals. Worms are usually soft-bodied, naked, and commonly without legs or other appendages; in short they may be any animal bearing even a superficial resemblance to the earthworm. The term now has no zoological standing but formerly was the group term for the old phylum Vermes The word worm is now employed most commonly as a part of a name, such as earthworm, flatworm, and roundworm.

The most common groups referred to as worms are Annelida (earthworms, sandworms, and clamworms), Platyhelminthes (flatworms, flukes, and tapeworms), and Nematoda (round and threadworms). There are several other groups where the term is employed as a part of the common name, including Rhynchocoela, ribbon worms; Chaetognatha, arrow worms; Acanthocephala, spiny-headed worms; and Nematomorpha, horsehair worms, or horsehair snakes.

The larvae of numerous insects are also called worms, for example, silkworms, wireworms, and armyworms *See* ACANTHOCEPHALA; ANNELIDA; ASCHELMINTHES; CHAETOGNATHA; PLATYHELMINTHES, RHYNCHOCOELA. [J.D.B]

Wren

Any member of the family Troglodytidae, a widely distributed songbird family, with 63 species in the Holarctic and Neotropical regions Ten occur in the United States

Wrens are usually small, compact, brown-streaked birds, with long tails and nervous, flitting habits. They are usually associated with shrubbery and underbrush The house wren, *Tryglodytes*

The house wren, Troglodytes aëdon; length 5¼ in (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

X Xenolith to Xylene

Xenolith

A foreign rock fragment caught up in solidifying lava or magma (rock melt). Xenoliths may be incorporated in abundance when magma intrudes shattered rocks. They may show a preferred or haphazard orientation (Fig. 1). Subsequent to incorporation they may recrystallize to hornfels or quartzite, and through reconstitution, new minerals may form. This metamorphic action is accomplished by heat and volatile constituents from the magma. By chemical reaction, xenoliths tend to be made over into minerals in equilibrium with the rock melt; and the magma becomes contaminated (Fig. 2). Some xenoliths appear only slightly affected; others are more or less rounded, digested, or disintegrated by the magma. See AUREOLE, CONTACT; IGNEOUS ROCKS; MAGMA. [C.A.CA.]

Fig. 1. *Xenoliths* in granodiorite, Sierra Nevada, Calif., appear soaked and drawn out parallel to the hammer handle. Hammer is 10 in. long. (USGS photo by W. B. Hamilton)

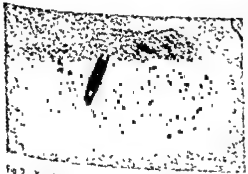
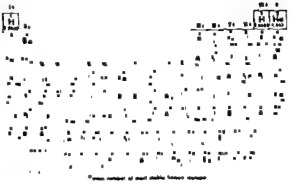


Fig. 2. Xenolith in granite, Sierra Nevada, California. Note sharper contact on one side and gradational contact against contaminated granite on other. (USGS photo by W. B. Hamilton)

Xenon

Chemical element number 54, xenon, Xe, is a member of group 0 of the periodic table. *See* INERT GASES.



Uses. Xenon is used to fill a type of flash bulb employed in photography, called electronic speed light. These bulbs produce a white light that has a good balance of all the colors in the visible spectrum, and can be used 10,000 times or more before burning out.

A xenon-filled arc lamp gives a light intensity approaching that of the carbon arc; it is particularly valuable in projecting motion pictures.

Xenon readily absorbs radiation, such as x-rays; when xenon is mixed with acetylene, for example, and treated with x-rays, the xenon absorbs the radiation and transmits it to the acetylene, resulting in polymerization of the acetylene.

Xenon has relatively high solubility in body cells. Experiments on humans have shown that breathing a mixture of 20% oxygen and 80% xenon rapidly produces deep anesthesia useful in surgery; the patients on whom this was tried awoke without unpleasant after-effects within 2 minutes of the time administration of the xenon was ended. Because the xenon is chemically inert, it cannot poison the patient, and there is no danger of fire or explosion as there is with ether, ethylene, or other flammable anesthetics. In another medical application, xenon injected into the skull gives clearer x-ray pictures than when air is injected; moreover, xenon did not give the patients the long-lasting headache that sometimes results when air is used in encephalography.

An important development in high-energy physics is the detection of nuclear radiation, such as γ-rays and mesons by bubble chambers, in which a liquid is kept at a temperature just above its boil-

Efforts were made to improve production, largely through attempts to simulate mechanically the manual effort of the puddler. All of these failed, economically or because the quality of the product was poor.

One method has been developed and has been in large-scale operation for a number of years. It is based upon application of technical principles related to wrought-iron structure, rather than mechanical manipulations. It is called the Aston process, and it is conducted in three main steps:

First, pig iron is refined by standard steel-making operations—the Bessemer converter is used.

Second, an iron silicate slag of desired composition is melted in a rotary furnace—iron ore and sand are the basic ingredients.

Third, the metal is poured in a steady stream into a large stationary vessel containing molten slag at a desired temperature below the freezing point of the iron, thus causing a sponge ball of wrought iron to solidify.

During instantaneous and progressive solidification of the metal, its dissolved gases are liberated with sufficient force to disintegrate the metal. The large volume of liquid slag is decanted and reused, and the sponge ball with its entrained slag is compacted into a large ingot while at a welding heat

The result is wrought iron of controlled quality. Unit masses and all mechanical equipment closely parallel those of steel making. Diversification of product is attained, which was not feasible with the limitations of the puddling process. *See* CAST IRON, IRON (EXTRACTION FROM ORE); IRON ALLOYS; STEEL. [J.A.]

Wulfenite

A mineral consisting of lead molybdate, $PbMoO_4$. W[illegible] occurs commonly in yellow orange red

granular. Its fracture is uneven Its hardness is 2.7–3 and its specific gravity 6.5–7 Its streak is white. It is easily fusible and is decomposed by hydrochloric or nitric acid with the separation of molybdic oxide.

Wulfenite occurs as a secondary mineral in the oxidized zone (of veins) of lead deposits associated with pyromorphite, cerussite, vanadinite, and other oxide zone minerals such as goethite and calcite

Wulfenite is found in numerous localities in the western and southwestern United States. Brilliant orange tabular wulfenite crystals up to 2 in. in size have been found from the Red Cloud and Hamburg mines in Yuma County, Arizona. *See* MOLYBDENUM. [E.C.T.C.]

Wurtz reaction

The reaction of an alkyl halide with an alkali metal to produce an alkane. Good yields are obtained in some cases (for example, in the preparation of 2.7-dimethyloctane from isopentyl bromide), but the formation of high-boiling by-products usually causes low yields of the desired alkanes

$$2RX + 2Na \rightarrow RR + 2NaX$$

$$3RX + 3R'X + 6Na \rightarrow RR + RR' + R'R' + 6NaX$$

The relative amounts of the three alkanes obtained by the reaction of the two different alkyl halides depends on the relative reactivities of the halides high yields of RR′ are obtained only when their reactivities are approximately the same. *See* ALKANE HALOGENATED HYDROCARBON; WURTZ-FITTIG REACTION. [L.S.]

Wurtz-Fittig reaction

A modification of the Wurtz reaction in which one of the participating halides is an aryl (aromatic) halide. In a typical example, equimolecular quantities of bromobenzene and butyl bromide are allowed to react at room temperature with slightly more than two equivalents of sodium metal to yield butylbenzene in 65–70% yield Other products are *n*-octane, $CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_3$, and biphenyl, $C_6H_5—C_6H_5$

$$C_6H_5Br + CH_3CH_2CH_2CH_2Br + 2Na \rightarrow C_6H_5CH_2CH_2CH_2CH_3 + 2NaBr$$

The reaction is most successful when the halides are bromides. Primary alkyl halides give better results than those obtained from secondary halides. Tertiary alkyl halides do not appear to be suitable.

It is characteristic of the Wurtz-Fittig reaction that the entering alkyl group does not undergo rearrangement. *See* AROMATIC HYDROCARBON; BENZENE; HALOGENATED HYDROCARBON; WURTZ REACTION [C.K.B.]

Wurtzilite

A black, infusible carbonaceous substance occurring in Uinta County, Utah It is insoluble in carbon disulfide, has a density of about 1.05, and consists of 79–80% carbon, 10.5–12.5% hydrogen, 4–6% sulfur, 1.8–2.2% nitrogen, and some oxygen.

Wurtzilite is derived from shale beds deposited near the close of Eocene (Green River) time. The material was introduced into the calcareous shale beds as a fluid after which it polymerized to form nodules or veins. *See* ALBERTITE, ELATERITE; IMPSONITE, see also ASPHALT AND ASPHALTITE [I.A.B.]

Wurtzite

A mineral with composition ZnS (zinc sulfide), crystallizing in the hexagonal system (dihexagonal pyramidal class). Crystals are pyramidal with the vertical axis polar. It exists most commonly in fibrous or columnar aggregates and banded crusts with resinous luster and brownish-black color. There is good prismatic cleavage, the hardness is 3½ and its specific gravity 4.0

Zinc sulfide is dimorphous, crystallizing as both wurtzite and sphalerite Wurtzite is the rarer and more unstable form and, unlike sphalerite, is rarely mined as an ore of zinc. It usually contains iron and cadmium, and a complete solid solution series extends to greenockite, CdS *See* CADMIUM, SPHALERITE. [C.S.HU.]

slender limbs and 5 toed feet with broad flat phalanges. They are restricted to the Paleocene deposits of Brazil and Argentina. There is only one family (Carodniidae) in the order and only one genus (Carodnia) in the family

The dentition is complete with strong, procumbent, chisel shaped *incisors*, *strong sharp-pointed canines*, and low-crowned cheek teeth with bilophodont molars:

$$\text{Incisors } \frac{3}{3}, \text{ canines } \frac{1}{1}, \text{ premolars } \frac{4}{4}, \text{ molars } \frac{3}{3}$$

Previously assigned to the order Pyrotheria, xenungulates appear to be most closely related to the Pyrotheria and the Dinocerata *See* DINOCERATA, PYROTHERIA [G.T.J.]

Xiphosura

A subclass of the class merostomata These "sword tailed" arthropods have three distinct body divisions: a prominent horseshoe-shaped, fused head and thorax, the prosoma; a segmented "abdomen," or opisthosoma; and a spikelike terminal segment, the telson. The Xiphosura includes three major subgroups, the Aglaspida, Synxiphosurida, and Limulida, which span over 500,000,000 years of evolution. All are, or were, inhabitants of shallow brackish water areas and tolerant of wide ranges of salinity

The Aglaspida were elongate Xiphosura with prominent compound eyes, free segments in the opisthosoma, and had a broad-based telson.

The Synxiphosurida usually lacked *compound* eyes. Some species had a distinct axial, or cardiac, region, free segments in the opisthosoma, and a broad, short telson

The Limulida include the species constituting a clearly recognizable horseshoe "crab" group: Belinuracea, Euproopacea, and Limulacea.

Distribution. The Mesozoic waters of Europe were the probable center of dispersal of the an-

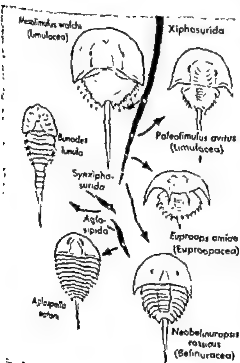


Fig 1 Phylogeny and representative species of Xiphosura

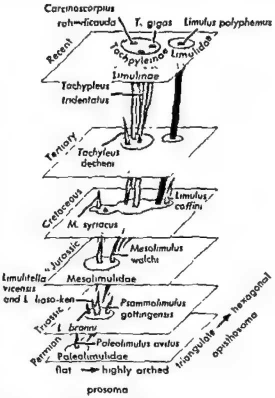


Fig 2. Limulacea family tree.

cestors of present species The geographical distribution of the Limulacea along the coasts of the United States is heaviest, and the

It is smaller to north and south, this indicates the Limulacea always have been tropicotemperate water species

The four living species of Limulacea have a restricted geographic range. *Limulus polyphemus*

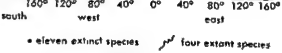


Fig 3 Geographical distribution of Limulacea, lower Permian to Recent

ing point. Nucleation by the radiation results in bubble formation along the path of the particle. The tracks made by the particles are then photographed. Liquid xenon is one of the liquids used in these bubble chambers

Xenon is used to fill neutron counters, x-ray counters, gas-filled thyratrons, and ionization chambers for cosmic rays; it is also used in high-pressure arc lamps for the production of ultraviolet radiation.

Between 3 and 5% of the fissions in a nuclear reactor using uranium as fuel lead to the formation of xenon-135. An interesting use for this xenon has been suggested. Because it is an effective neutron absorber, it might be purged out of the fuel zone by a stream of helium and passed into a confined space around the reactor to act as a neutron shield.

Occurrence. The only commercial source of xenon is the air. Traces of xenon are found in minerals and meteorites. Xenon constitutes 0 086 parts per million by volume of dry air, and this xenon is a mixture of the following isotopes, none of which is radioactive: 129 (26.44%), 131 (21.18%), 132 (26.89%), 134 (10 44%), 136 (8.87%), and 124, 126, 128, and 130 (6.18%) The relative abundances of the isotopes are given in parentheses.

A mixture of stable and radioactive xenon is formed in nuclear reactors in connection with the neutron fission of uranium This formation of xenon is a nuisance in the operation of the reactor, because xenon readily absorbs neutrons and therefore gradually "poisons" the reactor fuel by slowing down the fission process. The xenon must either be flushed out in a stream of gas such as helium, or the fuel must be removed from the reactor occasionally and treated chemically to purify the uranium.

It is estimated that about $3 \times 10^{-9}\%$ of the weight of the earth is xenon. Xenon also occurs outside the earth; the best estimate is that there are about 4 atoms of xenon per 1,000,000 atoms of silicon in the visible universe.

Discovery. Xenon was discovered in England by Sir William Ramsay and M W Travers in 1898. They separated it by fractional distillation from crude krypton, and identified it as a new element by discovering new lines in the emission spectrum of the residual gas

Properties. Xenon is colorless, odorless, and tasteless; it is a gas under ordinary conditions Some of its other properties are shown in the table.

Physical properties of xenon

Atomic number	54
[illegible]	131 30
[illegible]	−111 8
[illegible]	−108 1
[illegible] g/liter	5 8992
Liquid density at its boiling point, g/ml	3 06
Solubility in water, cm³ gas/1000 g water at 25°C and 1 atm pressure	101 5

Xenon does not form chemical compounds in the ordinary sense of the word, but it does form some weakly bonded clathrate compounds with substances such as water, hydroquinone, and phenol See CLATHRATE COMPOUNDS

Production and distribution. The commercial production of xenon is carried out in an air-separation plant. The air is liquefied and distilled. The oxygen is redistilled; the least volatile portion then contains small amounts of xenon and krypton which are adsorbed on silica gel from the liquid oxygen. The crude xenon and krypton thus obtained are separated and further purified by distillation and selective adsorption at controlled low temperatures on activated carbon The last of the impurities are removed by passing the xenon over hot titanium, which reacts with all but the inert gases.

Xenon could also be obtained from the gases produced in a nuclear reactor. First, the xenon would be separated from the other gases in the reactor, and then stored to eliminate radioactive xenon Because the longest half-life of any of the radioactive xenon isotopes produced in the nuclear reactor is only about 30 days, substantially all radioactivity would cease at the end of about 10 months The remaining stable xenon could then be purified. At present, the demand for xenon can easily be met by xenon produced from air

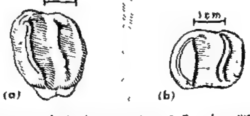
Xenon is sold at atmospheric pressure in sealed glass vessels, and under higher pressures in steel cylinders.

Analytical methods. The principal modern methods of detecting and quantitatively determining the xenon content in gases are mass spectrometry and gas chromatography Until these methods were developed, it was necessary to separate xenon from other inert gases by selective low-temperature adsorption on activated carbon in order to determine how much xenon was present in a mixture The older method of detecting xenon is by its characteristic emission spectrum, obtained by passing a gas sample through an electric discharge tube at low pressure and analyzing the light with a spectrometer See ATMOSPHERIC GASES, PRODUCTION OF, VAPOR LAMP [G.A.C.]

Bibliography F P Gross, Jr., Rare gases in everyday use, *J Chem Educ*, 18·533–539, 1941; R F. Kirk and D F Othmer (eds.), *Encyclopedia of Chemical Technology*, vol 7, 1951.

Xenungulata

Xenungulates are large, primitive, digitigrade, hoofed, tapirlike mammals, with relatively short,



(a) Upper and (b) lower molar of Carodnia vieirai from the Paleocene of Brazil. (After de Paula Couto, 1952)

(L.) is found along the Atlantic coast of North America from Nova Scotia to Yucatan. Three others, *Carcinoscorpius rotundicauda* (Latr.), *Tachypleus tridentatus* (Leach), and *Tachypleus gigas* (Müller), are endemic to the Indo-Pacific region.

Limulus polyphemus is the most studied species. Its common names include horsefoot or horseshoe "crab" (preferred), horsefeet, pan "crab," piggyback "crab," and unfortunately, "king crab," which is erroneous because *Paralithodes camtschatica* is a true king crab.

Life history. The life history of *Limulus* in Delaware Bay, where the largest population abounds, is characteristic of the species. Adults spend the colder months in 20- to 60-ft depths both within and outside the bay. Warming waters and increasing daylight stimulate the adults to migrate toward the sandy bay beaches to spawn. Most follow the path of least resistance and follow the strong flood tide currents into the eastern side of Delaware Bay. Fewer horseshoe "crabs" gather along the western shore, except when rough waves strike Cape May beaches during prevailing northwesterlies and drive the animals back into the deeper water. Then they may venture westward toward the wind protected shore.

On new moon, and especially full moon, high tides in May and June, thousands of the adults congregate on the beaches to spawn. Males "patrolling" along the shore grasp the females heading for the beach. In an excavation dug beneath her body, each female lays several thousand eggs which are fertilized when the spermatozoa are washed into the nests by the waves. After making one or more nests, the animals leave the beach when the tide ebbs. Limuli that do not beat the receding tide are stranded upon the tidal flats where they dig in to await the return of the next high water.

Minnows swarm along the beaches at night and feed upon the eggs washed out of the nests by wave action. During the day sharp eyed shore birds, mostly migrants such as plovers and sandpipers, dig into the nests and gorge themselves.

Horseshoe "crab" embryos begin to develop within the opaque, tough-coated eggs. In a few days each egg cover ... ent sphere ... balloons ... original e... all this transparent showcase, growth of the legs, their first feeble attempts to move, and then their almost incessant movement are clearly visible.

Within 2 weeks some of the tiny animals reach the larval stage and are ready to hatch although development and hatching may be delayed for several weeks. If the waves churn up a nest, the moving sand grains rupture the enclosing membrane and liberate the larvae. From then on, until they reach maturity in some 9–11 years and 15 or more stages later, their life is essentially that of seeking food and growing. Food consists mainly of marine worms such as *Nereis*, and soft shelled mollusks like *Mya arenaria*. The larger the immature animal, the more it moves toward the deeper bay water, from which, as an adult, it begins the spawning migration to create the life cycle anew.

The newly-hatched larva, the "trilobite" or Euproops stage, is in many respects still an embryo. Its incompletely formed digestive tract, nervous and circulatory systems develop further during early subsequent stages. The yolk-filled midgut serves as food for the larvae while they "learn" to right themselves from the upside-down position in which gravity held them during development in the egg. All ages of limuli swim on their backs, however, and usually at the water surface.

Molting. Molting occurs several times during the first 2 or 3 years of life, but rarely more than once thereafter until adult size is reached. It is preceded by the formation of the new skin, recognizable externally in the larger animals by a deeper olive-green coloration with yellow margins of the shell. The new skin forms pleats which unfold as the animal expands with the uptake of water during emergence. These pleats in the premolt skin give the hardened exoskeleton its dark-lined, mosaic appearance. A circumferential splitting, just within the ventral margin of the prosoma, is the first outward evidence of molting. Exuviation then proceeds slowly until the animal emerges one-third of the way and is loosened from the old shell. Emergence then proceeds more rapidly. The first-tailed stage emerges within an hour but an animal 100 mm wide may require over 24 hours. Younger specimens molt earlier than larger ones, with cast shells appearing in sequence during July and August on tidal flats and in beach wrack. Instinctively, the larger, annually molting animals seek the areas along the edges of tidal flats and dig in prior to molting. Burying protects them from their enemies, birds, fishes, and crabs. The moist substratum supplies water at all times.

Nutrition. In feeding, the pincer-tipped legs grasp and bring prey to the heavy bases of the "pushers," the fifth pair of legs, where shells of mollusks are crushed. A pair of paddlelike appendages and spines on the leg bases then push and "knead" the prey forward toward the mouth which is surrounded by the leg bases. Chelipeds in front of upper lip push food backwards. Chitinous ridges of the muscular gizzard, with the aid of ingested sand and shell, grind up prey into pulp which is squeezed into the midgut where it is partially digested. The resulting chyle is forced into the many-branched digestive gland where digestion within the cells completes the process. Large pieces of shell and pebbles are regurgitated from the gizzard, but fine sand may be passed with undigested residues in soft, slime-enveloped, fecal cylinders.

Anatomy. The carapace (test or shell) is composed of chitin, strengthened internally in the adults by bridgelike struts within the shell margins. Sensory structures on the carapace include

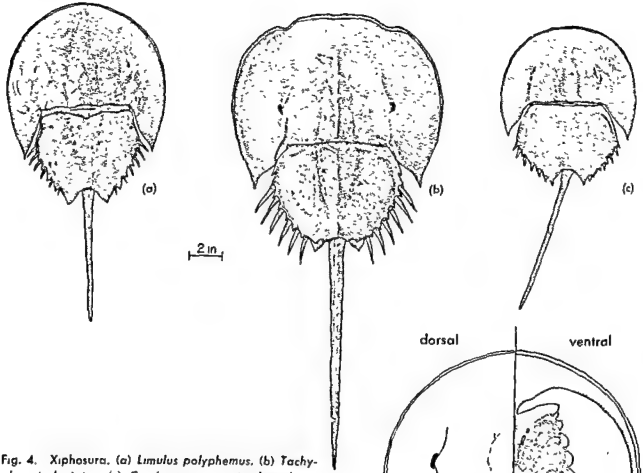


Fig. 4. Xiphosura. (a) *Limulus polyphemus.* (b) *Tachypleus tridentatus* (c) *Carcinoscorpius rotundicauda*

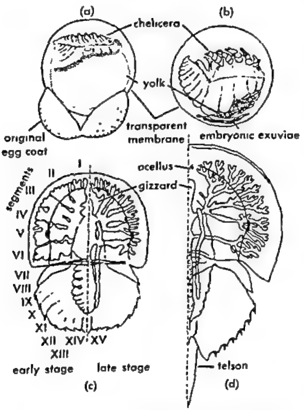


Fig. 5. Xiphosura. (a) Early embryo, side view. (b) Late embryo (c) Yolk-filled digestive tract of larva. Note segmentation of yolk in newly hatched stage and six lobes of digestive gland. (d) First-tailed stage.

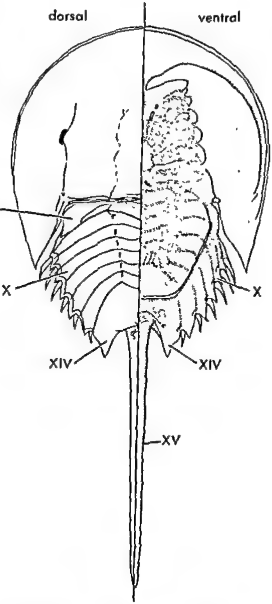


Fig 6 Segmentation. Inconclusive evidence indicates 15 segments (6 in prosoma, 9 in opisthosoma). The telson (XV) may be a terminal spine that has become articulated and has incorporated the anus into its base

Quartered animals are good eel bait and food for poultry and hogs. The roe reportedly increase growth of broilers and egg laying, although a fishy flavor is imparted to both flesh and eggs.

Limulus is a useful research animal It is relatively simple in structure, tenacious of life, and easy to obtain It is of interest as a living fossil

[C.N.S.]

Bibliography: F. A. Brown, Jr. (ed), *Selected Invertebrate Types*, 1950, R. C. Moore (ed.), *Treatise on Invertebrate Paleontology*, 1955; R. E. Snodgrass, *A Textbook of Arthropod Anatomy*, 1952

X-ray(s), physical nature of

X-rays, or roentgen rays, are electromagnetic waves in which periodically variable electric and magnetic fields are perpendicular to each other and to the direction of propagation. Thus they are identical in nature with visible light and all the other types of radiation that constitute the electromagnetic spectrum (ultraviolet, infrared, γ-rays from radioactive atomic disintegrations, microwaves, and radio or Hertzian waves). *See* ELECTROMAGNETIC RADIATION In general, x-rays are generated as the result of energy transitions of atomic electrons caused by the bombardment of a material of high atomic weight by high energy electrons

Röntgen's findings. The unequivocal establishment of the nature of these rays was not made in W. K. Röntgen's experiments following the discovery of "a new kind of ray" in 1895. In his first communication, Röntgen described the properties of these rays as follows. they were invisible; moved in straight lines; were unaffected by electric or magnetic fields, and hence not electrically charged; passed through matter opaque to ordinary light (since they penetrated through the black cardboard around his cathode-ray tube); were differentially absorbed by matter of different densities or of different atomic weights; affected photographic plates, produced fluorescence in certain chemicals, such as in the barium platinocyanide screen with which the initial discovery was made and in the wall of his glass tube opposite the cathode; produced ionization in gases; and were evidently produced by the stoppage at the anode of the beam of rays (identified by J J Thomson in 1897 as electrons) issuing from the cathode in his vacuum tube See CATHODE RAYS

Along with all these definitive characteristics of the rays, however, other crucial experiments designed to establish similarity or differences from ordinary light were clearly called for The fundamental optical properties of light were well established in 1895: reflection from mirrors, refraction in prisms (change in direction in passing from air into glass for example), by means of which a beam of white light could be spread out into a rainbow or spectrum of colors; diffraction by narrow slits or ruled gratings, also a method of producing spectra, and polarization, or constraint of the transverse vibrations to a single direction. In spite of the best efforts of Röntgen, no indubitable evidence of any of these four optical phenomena could be found. Hence the designation "x"—unknown—was assigned by Röntgen. Many theories were proposed to account for the apparently unique quality of x-rays, which seemed to be so closely similar and yet so greatly different from light; some suggestions were that they were vortex rings in the ether, and waves with longitudinal vibrations, that is, vibrations parallel to the direction of propagation as in sound waves, instead of transverse as with light.

Later discoveries. Inevitably, other scientists studying the enigma found the essential experimental conditions to prove that x-rays can be polarized (C. Barkla, 1905, by scattering from carbon); diffracted by crystals (M. von Laue, W. Friedrich, and P. Knipping, 1912); refracted in prisms and in crystals; reflected by mirrors; and diffracted by ruled gratings (A. Compton, 1921–1922). Instead of being refracted in passing from a less dense medium (air) to a more dense medium (a glass prism or a crystal) in the same direction as light so that the index of refraction is always greater than 1, x-rays are deviated in the opposite direction by a very small amount, so that the index of refraction is less than 1 by an amount as small as 10^{-6}. Thus total reflection from mirrors is observed only when the beam impinges at a very small grazing angle, a necessary condition understandably missed by Rontgen Similarly, the beam must graze a ruled diffraction grating if a spectrum is to be observed *See* X-RAY OPTICS

From 1895 to 1912 there seemed to be no analyzer capable of dispersing an x-ray beam into a spectrum The spectacular Laue diffraction pattern of a zinc sulfide crystal in 1912 proved the electromagnetic wave nature of x-rays and the ordered structure of crystals with atoms lying on families of planes to constitute three-dimensional diffraction gratings, all governed by the simple Bragg law

$$n\lambda = 2d \sin \theta$$

(which must be corrected for refraction in extremely accurate work) where n is an integer indicating the order of the spectrum, λ the wavelength, d the crystal lattice spacing of one set of planes, and θ the angle between the incident ray and this set of planes

The range of x-rays in the electromagnetic spectrum as excited in x-ray tubes by the bombardment of anode targets by cathode electrons under a high accelerating potential, overlaps the ultraviolet range on the order of 1000 angstroms (1 A is 10^{-8} cm) on the long wavelength side, and the shortest wavelength limit moves downward as voltages increase An accelerating potential of 10^9 volts, now readily generated, produces a λ of 0 00001 x 10^{-8} cm (10^{-5} A) An average wavelength used in research is 1 A, or about 1/6000 the wavelength of yellow light.

Quantum theory. In the consideration of roentgen

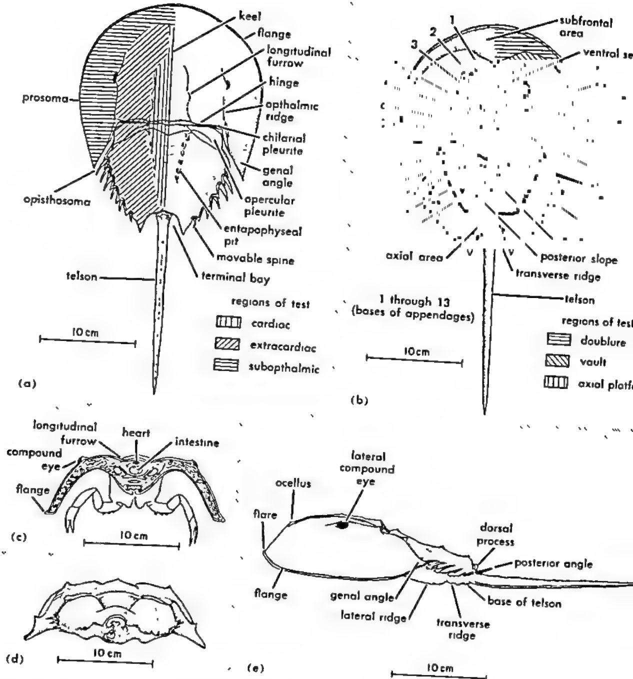


Fig. 7. Morphology of *Limulus*. (a) Dorsal view (b) Ventral view. (c) Cross section through prosoma. (d) Posterior aspect (e) Lateral view

chemoreceptors for taste on the bases of the legs and a pair of simple eyes, or ocelli The multilensed compound eyes form mosaic images and are sensitive to polarized light.

Adult males are easily distinguished from the larger females by the higher anterior arch of the prosoma, the pair of "fist-and-thumb" claspers used in mating, and by the hard, conical projections bearing the genital pores on the back of the operculum (eighth appendage) The female apertures are larger, soft, and oval.

The organ systems, except the peripheral blood vessels and nerves, digestive, and reproductive glands, which branch extensively into the lateral portions of the prosoma, are mainly within the prominent axial region. This region contains the musculature of the thirteen paired appendages, hinge, and telson, the digestive tube, the coxal (excretory) gland, the respiratory, and the central circulatory and nervous structures The dorsally situated tubular heart lies within a sinus which receives venous blood, the hemolymph, from the book gills, respiratory structures, and the body. The central nervous system and the major nerve branches are sheathed by arteries A plate of "cartilagenous" chitin, the endocranium, forms a roof over the brain. Flat, elliptical primitive white blood cells, the amoebocytes, circulate in the hemolymph When exposed to air the amoebocytes send out pseudopods and fragment, to form a clot. The blood pigment is hemocyanin. *See* RESPIRATORY PIGMENTS.

Importance. Horseshoe "crab" meal which has a protein content of 46%, is used in fertilizers

acterized by a vector normal to the plane and of a length inversely proportional to the d_{hkl}. Thus the set of parallel crystal planes characterized by d_{402} ($h = 4$, $k = 0$, $l = 2$) is represented by the point 402 in Fig. 1. The point of the reciprocal lattice which will be reciprocal to the set of planes is located on the normal drawn from the origin to the planes and at a distance d_{hkl}^* from the origin. The distance d_{hkl}^* is equal to a constant K divided by the spacing of the set of planes d_{hkl}. For theoretical work, it is convenient to make this constant unity.

For the practical purpose of interpreting experimental x-ray films, however, it is generally more convenient to introduce the wavelength λ of the radiation at this point and to take λ as the value for K. This is really only a scale factor; such reciprocal lattices are λ times the size of the reciprocal lattice used by the theoretician. Reciprocal lattice points corresponding to the various n orders of a given crystallographic set of planes are then placed along the same normal at n times the distance from the origin of the point corresponding to the set of planes of the first order. It can be shown that this procedure applied to a three-dimensional periodic lattice will result in a three-dimensional reciprocal lattice. Each point of coordinates hkl in this reciprocal lattice will be reciprocal to a set of planes of index hkl in the original lattice.

Sphere of reflection. The geometry of x-ray diffraction phenomena can be explained best in terms of a sphere of reflection (Fig. 2). The sphere is of unit radius, and it may be drawn to any convenient scale. The crystal C is imagined to be at the center of the sphere and the incident and transmitted x ray are assumed to travel along a diameter. The origin of the reciprocal lattice is placed at the point O where the transmitted beam emerges from the sphere of reflection. The reciprocal lattice, by virtue of the way it is defined, is tied to the actual lattice insofar as orientation is concerned. If the crystal (and therefore the lattice) is rotated about an axis of rotation, then the reciprocal lattice must be given a similar rotation about a parallel axis through its origin. In general, the space inside the sphere will contain many reciprocal lattice points, as will the space outside. However, if a reciprocal lattice point lies *on* the

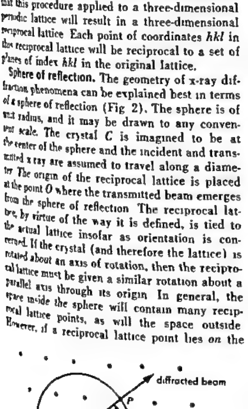


Fig. 2. Sphere of reflection. All the black dots represent reciprocal lattice points.

surface of the sphere of reflection, two important results follow: (1) the set of planes to which this point is the reciprocal obeys Bragg's law and reflects the incident x-ray beam; (2) the direction of the reflected ray will be from the point C to where the point P lies on the surface of the sphere.

This geometrical interpretation of Bragg's law enables one to understand readily both the geometry of reflection of x-rays and the manipulation of all the single-crystal cameras and apparatus now in use.

Recording techniques. The points of the reciprocal lattice may be considered as lying in plane layers which will in one way or another pass through the surface of the sphere of reflection. When they do, the crystallographic planes to which these points are reciprocal will reflect x-rays. In diffraction cameras using the moving film technique, by the use of opaque layer line screens, reflections from only one plane of reciprocal lattice points are permitted to reach the film. All other reflections are intercepted. If the film were not moving, these selected reflections would be confined to a line on the film; moving the film with the crystal by means of some appropriate coupling distributes the location of the points where the reflected rays hit the film into a plane and permits the unique and unequivocal indexing of the reflections.

In the case of the Schiebold-Sauter and Weissenberg techniques, this recording of the reflections due to the points of one layer of the reciprocal lattice results in a distorted transformation or picture of the reciprocal lattice layer, while in the deJong-Bouman and precession cameras, the layer of reciprocal lattice points is produced on the photographic film without geometric distortion.

In the early days of x-ray crystallography, the classical Bragg x-ray spectrometer was used with an ionization chamber to collect data on the intensity of x-ray reflections. It was soon found possible and more convenient to use photographic recording for such data collection. However, since about 1953, with the availability of highly stabilized x-ray diffraction units and sensitive, reliable detectors such as Geiger counters and scintillation detectors, and with the demand for increased accuracy in intensity measurements, there has been a return to direct recording.

Structural analysis. While the size and shape of the unit cell determines the geometry of the x-ray reflections, the intensity of the x-ray reflections is determined by the number, character, and distribution of the atoms within the unit cell. Consequently it is this intensity information which must be used to work out the crystal structure. The first crystal structures to be determined were, of course, the relatively simple ones. The structures were determined in a direct but very unsophisticated way; however, the results of these early structural studies provided a basis for the science of crystal chemistry. Information about the size of atoms, the kind of environments they favor, the nature of the bonds between atoms, and the angles between bonds was obtained. This information could then be used in

accordance with the laws first enunciated by M. Planck and extended by A. Einstein early in the twentieth century. In diffraction, refraction, polarization, and interference phenomena, x-rays, together with all other related radiations, appear to act as waves and λ has a real significance. Beams of corpuscular electrons and neutrons are diffracted so that they too have wavelengths (*see* ELECTRON DIFFRACTION; NEUTRON DIFFRACTION). In other phenomena, such as the appearance of sharp spectral lines, a definite short wavelength limit λ_0 of the continuous "white" spectrum [defined by $\lambda_0 = hc/eV$, where h is Planck's constant, c the velocity of electromagnetic radiation (including light and x-rays), e the charge of the electron, and V the accelerating voltage], the shift in wavelength of x-rays scattered by electrons in atoms (Compton effect), and the photoelectric effect, the energy seems to be propagated and transferred in quanta defined by values of $h\nu$, where the frequency ν is c/λ. These quanta are called photons. *See* COMPTON EFFECT; PHOTOEMISSION; QUANTUM MECHANICS.

Applications. Important uses have been found for x-rays in many fields of scientific endeavor. For example, roentgen spectrometry is the science of measuring λ values with a known crystal of lattice spacing d; roentgen diffractometry is the science of determining unknown values of d, and thereby crystal structures, with x-ray beams of known λ. In both cases, the experimental measurement is that of the angle θ. Extensive tables of the wavelengths of x-ray emission lines in series (K, L, M, and so on) and so-called absorption edges, characteristic of the chemical elements, afford the necessary information for chemical analyses, exactly as in the case of optical emission spectra, and for derivation of theories of atomic structure to account for the origin of spectra. *See* X-RAY CRYSTALLOGRAPHY; X-RAY DIFFRACTION; X-RAY FLUORESCENCE ANALYSIS; X-RAY POWDER METHODS; *see also* HISTORADIOGRAPHY; MICROSCOPE, X-RAY; MICRORADIOGRAPHY; RADIATION BIOLOGY; RADIOGRAPHY; RADIOLOGY; X-RAY TUBE. [G.L.CL.]

Bibliography: G. L. Clark, *Applied X-rays*, 4th ed., 1955; A. H. Compton and S. K. Allison, 2d ed., *X-rays in Theory and Experiment*, 1935; K. Lonsdale, *Crystals and X-rays*, 1949; F. K. Richtmyer, E. H. Kennard, and T. Lauritsen, *Introduction to Modern Physics*, 5th ed., 1955

X-ray crystallography

The study of crystal structure by x-ray diffraction techniques. In 1912, the German physicist Max von Laue suggested that crystals might be employed as natural diffraction gratings in the study of x-rays. The experiment was successful, and provided a powerful tool for investigating both the structure of crystals and the nature of x-rays. This article is concerned with the former application. For the theoretical and experimental aspects of x-ray diffraction, *see* X-RAY DIFFRACTION.

Structurally, a crystal is a three-dimensional periodic arrangement in space of atoms, groups of atoms, or molecules. If the periodicity of this pattern extends throughout a given piece of material, one speaks of a single crystal. The exact structure of any given crystal is determined if the locations of all atoms making up the three-dimensional periodic pattern called the unit cell are known (*see* CRYSTALLOGRAPHY; CRYSTAL STRUCTURE). The very close and periodic arrangement of the atoms in a crystal permit it to act as a diffraction grating for x-rays. W. H. Bragg treated the phenomenon of the interference of x-rays with crystals as if the x-rays were being reflected by parallel equidistant planes of lattice points. He used the integer n to indicate the order of the interference. Bragg's equation relating the spacing d of planes, glancing angle θ, and wavelength λ, is

$$n\lambda = 2d_{hkl} \sin \theta$$

Modern crystallographers almost never refer to the nth order of an hkl reflection. Instead they use indices which contain the common factor n. For a discussion of Bragg's relation, *see* X-RAY POWDER METHODS.

For a crystal with known unit-cell size and shape, and with arbitrary orientation in a parallel beam of x-rays of known wavelength, several questions must be answered. First, it must be ascertained what plane, if any, is obeying Bragg's law and is reflecting the rays. Second, the direction of the reflected x-ray must be accurately measured. The Bragg law treatment does not permit one to answer these questions readily, and for this reason the concept of a reciprocal lattice model was introduced.

Reciprocal lattice. Figure 1 is a sketch of a simple monoclinic unit crystal cell. A lattice, reciprocal to a three-dimensional lattice, is constructed in the following way. An origin is chosen and is given the coordinates 000. Through this origin, perpendiculars to the different parallel sets of crystallographic planes of lattice points are imagined as drawn. For a given size and shape of unit cell, each of these sets of crystal planes of index hkl is char-

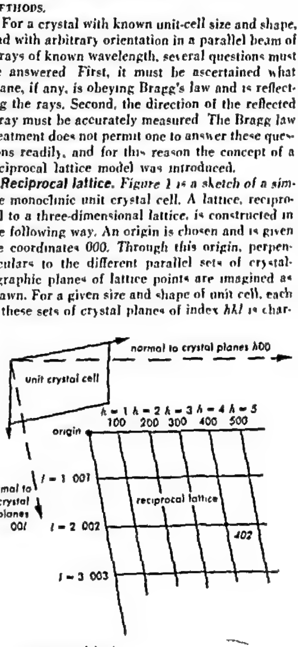


Fig. 1 Reciprocal lattice.

acterized by a vector normal to the plane and of a length inversely proportional to the d_{hkl}. Thus the set of parallel crystal planes characterized by d_{402} ($h = 4$, $k = 0$, $l = 2$) is represented by the point 402 in Fig. 1 The point of the reciprocal lattice which will be reciprocal to the set of planes is located on the normal drawn from the origin to the planes and at a distance d_{hkl}^* from the origin. The distance d_{hkl}^* is equal to a constant K divided by the spacing of the set of planes d_{hkl}. For theoretical work, it is convenient to make this constant unity

For the practical purpose of interpreting experimental x-ray films however, it is generally more convenient to introduce the wavelength λ of the radiation at this point and to take λ as the value for K This is really only a scale factor; such reciprocal lattices are λ times the size of the reciprocal lattice used by the theoretician. Reciprocal lattice points corresponding to the various n orders of a given crystallographic set of planes are then placed along the same normal at n times the distance from the origin of the point corresponding to the set of planes of the first order It can be shown that this procedure applied to a three-dimensional periodic lattice will result in a three-dimensional reciprocal lattice Each point of coordinates hkl in this reciprocal lattice will be reciprocal to a set of planes of index hkl in the original lattice

Sphere of reflection. The geometry of x-ray diffraction phenomena can be explained best in terms of a sphere of reflection (Fig 2). The sphere is of unit radius, and it may be drawn to any convenient scale The crystal C is imagined to be at the center of the sphere and the incident and transmitted x ray are assumed to travel along a diameter The origin of the reciprocal lattice is placed at the point O where the transmitted beam emerges from the sphere of reflection. The reciprocal lattice, by virtue of the way it is defined, is tied to the actual lattice insofar as orientation is concerned. If the crystal (and therefore the lattice) is rotated about an axis of rotation, then the reciprocal lattice must be given a similar rotation about a parallel axis through its origin In general, the space inside the sphere will contain many reciprocal lattice points, as will the space outside. However, if a reciprocal lattice point lies on the

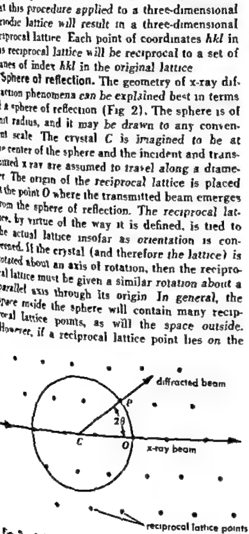


Fig 2 Sphere of reflection All the black dots represent reciprocal lattice points.

surface of the sphere of reflection, two important results follow: (1) the set of planes to which this point is the reciprocal obeys Bragg's law and reflects the incident x-ray beam; (2) the direction of the reflected ray will be from the point C to where the point P lies on the surface of the sphere.

This geometrical interpretation of Bragg's law enables one to understand readily both the geometry of reflection of x-rays and the manipulation of all the single-crystal cameras and apparatus now in use.

Recording techniques. The points of the reciprocal lattice may be considered as lying in plane layers which will in one way or another pass through the surface of the sphere of reflection. When they do, the crystallographic planes to which these points are reciprocal will reflect x-rays. In diffraction cameras using the moving film technique, by the use of opaque layer line screens, reflections from only one plane of reciprocal lattice points are permitted to reach the film. All other reflections are intercepted If the film were not moving, these selected reflections would be confined to a line on the film; moving the film with the crystal by means of some appropriate coupling distributes the location of the points where the reflected rays hit the film into a plane and permits the unique and unequivocal indexing of the reflections.

In the case of the Schiebold-Sauter and Weissenberg techniques, this recording of the reflections due to the points of one layer of the reciprocal lattice results in a distorted transformation or picture of the reciprocal lattice layer, while in the deJong-Bouman and precession cameras, the layer of reciprocal lattice points is produced on the photographic film without geometric distortion.

In the early days of x-ray crystallography, the classical Bragg x-ray spectrometer was used with an ionization chamber to collect data on the intensity of x-ray reflections. It was soon found possible and more convenient to use photographic recording for such data collection. However, since about 1953, with the availability of highly stabilized x-ray diffraction units and sensitive, reliable detectors such as Geiger counters and scintillation detectors, and with the demand for increased accuracy in intensity measurements, there has been a return to direct recording.

Structural analysis. While the size and shape of the unit cell determines the geometry of the x-ray reflections, the intensity of the x-ray reflections is determined by the number, character, and distribution of the atoms within the unit cell. Consequently it is this intensity information which must be used to work out the crystal structure The first crystal structures to be determined were, of course, the relatively simple ones The structures were determined in a direct but very unsophisticated way, however, the results of these early structural studies provided a basis for the science of crystal chemistry Information about the size of atoms, the kind of environments they favor, the nature of the bonds between atoms, and the angles between bonds was obtained. This information could then be used

working out the more complicated structures. Soon, a point of structural complexity was reached where *primitive and unsophisticated* methods were inadequate. The use of Fourier series analysis to represent the crystal structure was introduced. In principle, if the crystal structure (that is, the type and arrangement of the atoms in each cell) is known, computation of the intensity of the different x-ray reflections to be expected is comparatively straightforward.

A structure is deemed good if the comparison between observed intensity I and computed intensity is satisfactory. Obviously, the more accurate the intensity data, the more reliable is the structure derivable from the data. The correlation factor R generally used is given by

$$R = \frac{\Sigma(|F_{calc}| - |F_{obs}|)}{\Sigma|F_{calc}|}$$

where $F \propto \sqrt{I}$. The smaller this value, the more reliable the structure. With modern methods of data collection and structure determination, R values of 10% for comparatively complicated structures are not uncommon.

It is apparent that there can be no simple direct approach to the computation of a crystal structure. Coefficients of the Fourier series, representing the crystal structure, are F_{hkl} values with appropriate associated phase angles (that is, complex numbers), while it appears that only the magnitude of F_{hkl} can be obtained from x-ray observations. Much of the development in x-ray crystallography since about 1930 has been devoted to the problem of how to get around this difficulty. If an approximately correct structure can be arrived at in one way or another, then the method of Fourier series lends itself to a successive approximation approach to as complete a structure determination as the data warrant. The determination of a crystal structure depends upon the ability to derive an acceptable first approximate structure.

One of the best ways to establish a structure is to know enough about the way atoms and molecules usually pack in crystals to permit making a sophisticated guess at an approximate structure. The quality of the guess can then be tested by applying the method of successive approximations, using Fourier methods. If the structure refines to an acceptable R value, then an acceptable guess has been made. However, only rarely can a workable guess be made; other methods must be used to arrive at this first approximation. Among these is the Patterson approach, in which one uses observables, namely, $|F_{hkl}|^2$, as coefficients of a Fourier series. The maxima in this summation then correspond not to atomic centers but rather to the ends of interatomic vectors. In other words, the Patterson series describes a distribution function. When heavy atoms are present, the Patterson series summation can often be interpreted in terms of the location of the heavy atoms. This knowledge is often sufficient to enable the method of successive approximations to be used.

Another approach involves the use of isomorphous substitution. If crystals of two substances can be obtained that differ only in the nature of one of their atoms and that crystallize in the same crystal structure, then the differences in their x-ray reflections can be used to start the approximation procedure. Statistical analyses of the x-ray data are also used to obtain a starting point. All of these methods involving Fourier series require extensive computation.

More recently, the phenomenon of anomalous dispersion has been used to determine these missing phase angles. This approach is successful only when the crystal structure does not possess a center of symmetry and is based essentially on the fact that when radiation of wavelengths slightly shorter than the characteristic absorption edge of one of the constituent atoms is used, then the intensity of the reflection of x-rays from one side of a set of crystal planes is different from that of the other side. J. M. Bijvoet and R. Pepinsky have used this approach to work out the absolute configuration of molecules possessing asymmetrical carbon atoms. For additional information on anomalous dispersion, *see* ABSORPTION (ELECTROMAGNETIC RADIATION). For a discussion of absorption edges, *see* X-RAY FLUORESCENCE ANALYSIS. [I.F.]

Bibliography: W. H. Bragg (ed.), *The Crystalline State*, 1933–1953; M. J. Buerger, *X-Ray Crystallography*, 1942; M. J. Buerger, *Photography of the Reciprocal Lattice*, ACA Monograph 1, 1944; C. W. Bunn, *Chemical Crystallography*, 1945.

X-ray diffraction

The scattering of x-rays by matter with accompanying variation in intensity in different directions due to interference effects. X-ray diffraction is one of the most important tools of solid-state physics, since it constitutes a powerful and readily available method for determining atomic arrangements in matter. X-ray diffraction methods depend upon the fact that x-ray wavelengths of the order of 1 angstrom (A) are readily available and that this is the order of magnitude of atomic dimensions. When an x-ray beam falls on matter, scattered x-radiation is produced by all the atoms. These scattered waves spread out spherically from all the atoms in the sample, and the interference effects between the scattered radiation from the different atoms cause the intensity of the scattered radiation to exhibit maxima and minima in various directions. *See* DIFFRACTION.

Some of the uses of x-ray diffraction are (1) differentiation between crystalline and amorphous materials; (2) determination of the structure of crystalline materials (crystal axes, size and shape of the unit cell, positions of the atoms in the unit cell); (3) determination of electron distribution within the atoms, and throughout the unit cell; (4) determination of the orientation of single crystals; (5) determination of the texture of polygrained materials; (6) identification of crystalline phases and measurement of the relative propor-

tions, (7) measurement of limits of solid solubility, and determination of phase diagrams; (8) measurement of strain and small grain size; (9) measurement of various kinds of randomness, disorder, and imperfections in crystals; (10) determination of radial distribution functions for amorphous solids and liquids

For the study of crystal structure by x-ray diffraction techniques *see* X-RAY CRYSTALLOGRAPHY.

DIFFRACTION THEORY

When x rays fall on the atoms of a substance, the scattered radiation is of two kinds: Compton modified scattering of increased wavelength which is incoherent with respect to the primary beam (*see* COMPTON EFFECT), and unmodified scattering coherent with the primary beam Because of interference effects from the unmodified scattering by the different atoms of the sample, the intensity of unmodified scattering varies in different directions. A diagram of this variation in direction of intensity of unmodified scattering is called the diffraction pattern of the substance This pattern is determined by the kinds of atoms and their arrangement in the sample; for simple structures the atomic arrangement is readily deduced from the diffraction pattern

The atomic scattering factor f is defined as the ratio of the amplitude of unmodified scattering by an atom to the amplitude of scattering by a free electron wh[illegible]

[illegible] λ the wavelength, from an initial value $f = Z$, where Z is the number of electrons in the atom However, if the x ray wavelength is close to an absorption edge of the atom, f becomes complex. (For a definition of absorption edges, see X-RAY FLUORESCENCE ANALYSIS.) If the electron density in the atom has spherical symmetry and if the x-ray wavelength is small compared to all the absorption edge wavelengths,

$$f = \int_0^\infty 4\pi r^2 \rho(r) \frac{\sin kr}{kr}\, dr \qquad (1)$$

where $k = 4\pi(\sin\theta)/\lambda$ and $\rho(r)$ is the electron density (electrons per unit volume).

A crystalline structure is one in which a unit of structure called the unit cell repeats at regular intervals in three dimensions The repetition in space is determined by three noncoplanar vectors $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$, called the crystal axes. The positions of the atoms in the unit cell are expressed by a set of base vectors $\mathbf{r}_n$. The position of atom n in the unit cell $q_1q_2q_3$ is given by

$$\mathbf{R}_{nq} = q_1\mathbf{a}_1 + q_2\mathbf{a}_2 + q_3\mathbf{a}_3 + \mathbf{r}_n$$

See CRYSTALLOGRAPHY; CRYSTAL STRUCTURE.

For a crystal containing $N_1N_2N_3$ repetitions in the $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$ directions, the intensity of unmodified scattering is given by

$$\begin{aligned} I &= I_e \sum_{nq} f_n \exp\left[\frac{2\pi i}{\lambda}(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{R}_{nq}\right] \sum_{n'q'} f_{n'}^* \exp\left[\frac{-2\pi i}{\lambda}(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{R}_{n'q'}\right] \\ &= I_e FF^* \frac{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot N_1\mathbf{a}_1]}{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_1]} \frac{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot N_2\mathbf{a}_2]}{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_2]} \frac{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot N_3\mathbf{a}_3]}{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_3]} \end{aligned} \qquad (2)$$

Here I_e is the intensity, at a distance R and angle 2θ, scattered by a free electron according to classical theory, and $FF^* = |F|^2$. For an unpolarized primary beam of intensity I_o,

$$I_e = I_o \frac{e^4}{m^2c^4R^2}\left(\frac{1+\cos^2 2\theta}{2}\right) \qquad (3)$$

and $e^4/(m^2c^4) = 7.94 \times 10^{-26}$ cm^2 if R is expressed in cm Here m is the mass of the electron, c is the velocity of light, and F is the structure factor, a complex quantity given by a summation over all the atoms of the unit cell

$$F = \sum_n f_n \exp\left[\frac{2\pi i}{\lambda}(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{r}_n\right] \qquad (4)$$

where $\mathbf{s}_0$ and $\mathbf{s}$ are unit vectors in the directions of the primary and diffracted beams.

Laue equations and Bragg's law. The condition for a crystalline reflection is that the three quotients of Eq (2) exhibit maxima, and this occurs if all three denominators vanish. Expressing the denominators in terms of three integers α, β, and γ, the three Laue equations are obtained. These express the condition for a diffracted beam

$$\begin{aligned} (\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_1 &= \alpha\lambda \\ (\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_2 &= \beta\lambda \\ (\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_3 &= \gamma\lambda \end{aligned} \qquad (5)$$

It is convenient to introduce the concept of sets of crystallographic planes As illustrated by Fig. 1, the set of planes with Miller indices hkl is a set

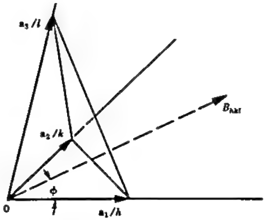


Fig 1 Crystallographic planes with Miller indices *hkl*

working out the more complicated structures. Soon, a point of structural complexity was reached where primitive and unsophisticated methods were inadequate. The use of Fourier series analysis to represent the crystal structure was introduced. In principle, if the crystal structure (that is, the type and arrangement of the atoms in each cell) is known, computation of the intensity of the different x-ray reflections to be expected is comparatively straightforward.

A structure is deemed good if the comparison between observed intensity I and computed intensity is satisfactory. Obviously, the more accurate the intensity data, the more reliable is the structure derivable from the data. The correlation factor R generally used is given by

$$R = \frac{\Sigma(|F_{calc}| - |F_{obs}|)}{\Sigma|F_{calc}|}$$

where $F \propto \sqrt{I}$. The smaller this value, the more reliable the structure. With modern methods of data collection and structure determination, R values of 10% for comparatively complicated structures are not uncommon.

It is apparent that there can be no simple direct approach to the computation of a crystal structure. Coefficients of the Fourier series, representing the crystal structure, are F_{hkl} values with appropriate associated phase angles (that is, complex numbers), while it appears that only the magnitude of F_{hkl} can be obtained from x-ray observations. Much of the development in x-ray crystallography since about 1930 has been devoted to the problem of how to get around this difficulty. If an approximately correct structure can be arrived at in one way or another, then the method of Fourier series lends itself to a successive approximation approach to as complete a structure determination as the data warrant. The determination of a crystal structure depends upon the ability to derive an acceptable first approximate structure.

One of the best ways to establish a structure is to know enough about the way atoms and molecules usually pack in crystals to permit making a sophisticated guess at an approximate structure. The quality of the guess can then be tested by applying the method of successive approximations, using Fourier methods. If the structure refines to an acceptable R value, then an acceptable guess has been made. However, only rarely can a workable guess be made; other methods must be used to arrive at this first approximation. Among these is the Patterson approach, in which one uses observables, namely, $|F_{hkl}|^2$, as coefficients of a Fourier series. The maxima in this summation then correspond not to atomic centers but rather to the ends of interatomic vectors. In other words, the Patterson series describes a distribution function. When heavy atoms are present, the Patterson series summation can often be interpreted in terms of the location of the heavy atoms. This knowledge is often sufficient to enable the method of successive approximations to be used.

Another approach involves the use of isomorphous substitution. If crystals of two substances can be obtained that differ only in the nature of one of their atoms and that crystallize in the same crystal structure, then the differences in their x-ray reflections can be used to start the approximation procedure. Statistical analyses of the x-ray data are also used to obtain a starting point. All of these methods involving Fourier series require extensive computation.

More recently, the phenomenon of anomalous dispersion has been used to determine these missing phase angles. This approach is successful only when the crystal structure does not possess a center of symmetry and is based essentially on the fact that when radiation of wavelengths slightly shorter than the characteristic absorption edge of one of the constituent atoms is used, then the intensity of the reflection of x-rays from one side of a set of crystal planes is different from that of the other side. J. M. Bijvoet and R. Pepinsky have used this approach to work out the absolute configuration of molecules possessing asymmetrical carbon atoms. For additional information on anomalous dispersion, *see* ABSORPTION (ELECTROMAGNETIC RADIATION). For a discussion of absorption edges, *see* X-RAY FLUORESCENCE ANALYSIS. [I.F.]

Bibliography: W. H. Bragg (ed.), *The Crystalline State*, 1933–1953; M. J. Buerger, *X-Ray Crystallography*, 1942; M. J. Buerger, *Photography of the Reciprocal Lattice*, ACA Monograph 1, 1944; C. W. Bunn, *Chemical Crystallography*, 1945.

X-ray diffraction

The scattering of x-rays by matter with accompanying variation in intensity in different directions due to interference effects. X-ray diffraction is one of the most important tools of solid-state physics, since it constitutes a powerful and readily available method for determining atomic arrangements in matter. X-ray diffraction methods depend upon the fact that x-ray wavelengths of the order of 1 angstrom (A) are readily available and that this is the order of magnitude of atomic dimensions. When an x-ray beam falls on matter, scattered x-radiation is produced by all the atoms. These scattered waves spread out spherically from all the atoms in the sample, and the interference effects between the scattered radiation from the different atoms cause the intensity of the scattered radiation to exhibit maxima and minima in various directions. *See* DIFFRACTION.

Some of the uses of x-ray diffraction are (1) differentiation between crystalline and amorphous materials; (2) determination of the structure of crystalline materials (crystal axes, size and shape of the unit cell, positions of the atoms in the unit cell); (3) determination of electron distribution within the atoms, and throughout the unit cell; (4) determination of the orientation of single crystals; (5) determination of the texture of polygrained materials; (6) identification of crystalline phases and measurement of the relative propor-

tions, (7) measurement of limits of solid solubility, and determination of phase diagrams; (8) measurement of strain and small grain size; (9) measurement of various kinds of randomness, disorder, and imperfections in crystals; (10) determination of radial distribution functions for amorphous solids and liquids

For the study of crystal structure by x-ray diffraction techniques *see* X-RAY CRYSTALLOGRAPHY.

DIFFRACTION THEORY

When x rays fall on the atoms of a substance, the scattered radiation is of two kinds: Compton modified scattering of increased wavelength which is incoherent with respect to the primary beam (*see* COMPTON EFFECT), and unmodified scattering coherent with the primary beam Because of interference effects from the unmodified scattering by the different atoms of the sample, the intensity of unmodified scattering varies in different directions. A diagram of this variation in direction of intensity of unmodified scattering is called the diffraction pattern of the substance This pattern is determined by the kinds of atoms and their arrangement in the sample; for simple structures the atomic arrangement is readily deduced from the diffraction pattern.

The atomic scattering factor f is defined as the

electron which scatters according to classical theory In general f is a real number which decreases with $(\sin\theta)/\lambda$, where θ is *the grazing angle* and λ the wavelength, from an initial value $f = Z$, where Z is the number of electrons in the atom However, if the x-ray wavelength is close to an absorption edge of the atom, f becomes complex (For a definition of *absorption edges*, see X-RAY FLUORESCENCE ANALYSIS) If the electron density in the atom has spherical symmetry and if the x-ray wavelength is small compared to all the absorption edge wavelengths,

$$f = \int_0^\infty 4\pi r^2 \rho(r) \frac{\sin kr}{kr}\, dr \qquad (1)$$

where $k = 4\pi(\sin\theta)/\lambda$ and $\rho(r)$ is the electron density (electrons per unit volume).

A crystalline structure is one in which a unit of structure called the unit cell repeats at regular intervals in three dimensions. The repetition in space is determined by three noncoplanar vectors $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$, called the crystal axes The *positions* of the atoms in the unit cell are expressed by a set of base vectors $\mathbf{r}_n$. The position of atom n in the unit cell $q_1q_2q_3$ is given by

$$\mathbf{R}_{nq} = q_1\mathbf{a}_1 + q_2\mathbf{a}_2 + q_3\mathbf{a}_3 + \mathbf{r}_n$$

See CRYSTALLOGRAPHY, CRYSTAL STRUCTURE.

For a crystal containing $N_1N_2N_3$ repetitions in the $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$ directions, the intensity of unmodified scattering is given by

$$I = I_e \sum_{nq} f_n \exp\left[\frac{2\pi i}{\lambda}(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{R}_{nq}\right] \sum_{n'q'} f_{n'} \exp\left[\frac{-2\pi i}{\lambda}(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{R}_{n'q'}\right]$$
$$= I_e FF^* \frac{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot N_1\mathbf{a}_1]}{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_1]} \frac{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot N_2\mathbf{a}_2]}{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_2]} \frac{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot N_3\mathbf{a}_3]}{\sin^2[(\pi/\lambda)(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_3]} \qquad (2)$$

Here I_e is the intensity, at a distance R and angle 2θ, scattered by a free electron according to classical theory, and $FF^* = |F|^2$. For an unpolarized primary beam of intensity I_o,

$$I_e = I_o \frac{e^4}{m^2c^4R^2}\left(\frac{1+\cos^2 2\theta}{2}\right) \qquad (3)$$

and $e^4/(m^2c^4) = 7.94 \times 10^{-26}$ cm² if R is expressed in cm. Here m is the mass of the electron. c is the velocity of light, and F is the structure factor. a complex quantity given by a summation over all the atoms of the unit cell

$$F = \sum_n f_n \exp\left[\frac{2\pi i}{\lambda}(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{r}_n\right] \qquad (4)$$

where $\mathbf{s}_0$ and $\mathbf{s}$ are unit vectors in the directions of the primary and diffracted beams.

Laue equations and Bragg's law. The condition for a crystalline reflection is that the three quotients of Eq (2) exhibit maxima, and this occurs if all three denominators vanish. Expressing the denominators in terms of three integers α, β, and γ, the three Laue equations are obtained. These express the condition for a diffracted beam

$$\begin{aligned}(\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_1 &= \alpha\lambda\\ (\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_2 &= \beta\lambda\\ (\mathbf{s}-\mathbf{s}_0)\cdot\mathbf{a}_3 &= \gamma\lambda\end{aligned} \qquad (5)$$

It is convenient to introduce the concept of sets of crystallographic planes As illustrated by Fig. 1, the set of planes with Miller indices hkl is a set

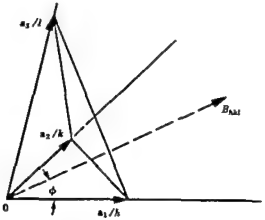


Fig 1. Crystallographic planes with Miller indices *hkl*

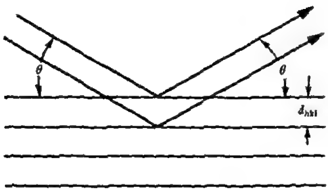


Fig. 2. Interference conditions involved in the Bragg law.

of parallel equidistant planes, one of which passes through the origin, and the next nearest makes intercepts a_1/h, a_2/k, a_3/l on the three crystallographic axes.

In terms of sets of planes hkl, the diffraction conditions are expressed by the Bragg law

$$\lambda = 2d_{hkl} \sin \theta \tag{6}$$

where θ is the angle which the primary and diffracted beams make with the planes hkl, and d_{hkl} is the spacing of the set. As seen from Fig. 2, the Bragg law is simply the condition that the path difference for rays diffracted from two successive hkl planes shall be one wavelength. In the early days of x-ray diffraction, the Bragg law was written $n\lambda = 2d \sin \theta$, and $n = 1, 2, 3$ corresponded to first-, second-, and third-order diffraction from the planes of spacing d. That notation has been largely dropped, and instead of being called second-order diffraction from planes hkl, it is called diffraction from the planes $2h, 2k, 2l$. For an extended discussion of the Bragg law, *see* X-RAY POWDER METHODS.

Reciprocal lattice. The understanding and interpretation of x-ray diffraction in crystals is greatly facilitated by the concept of a reciprocal lattice. In terms of the crystal axes $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$, three reciprocal vectors are defined by

$$\mathbf{b}_1 = \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3} \qquad \mathbf{b}_2 = \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3}$$
$$\mathbf{b}_3 = \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3} \tag{7}$$

From these definitions it follows that

$$\mathbf{a}_i \cdot \mathbf{b}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{8}$$

In terms of integers hkl, the terminal points of the vectors

$$\mathbf{B}_{hkl} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3 \tag{9}$$

generate a lattice of points called the reciprocal lattice. Each point in the lattice is specified by the integers hkl, and the vectors $\mathbf{B}_{hkl}$ represent two important properties of the sets of hkl planes: (1) $\mathbf{B}_{hkl}$ is perpendicular to the hkl planes; (2) $|\mathbf{B}_{hkl}| = 1/d_{hkl}$. These two relations are readily proved from the geometry of Fig. 1. As seen, $\mathbf{a}_2/k - \mathbf{a}_1/h$ and $\mathbf{a}_3/l - \mathbf{a}_2/k$ are vectors lying in the hkl plane. From Eqs. (8) and (9),

$$\left(\frac{\mathbf{a}_2}{k} - \frac{\mathbf{a}_1}{h}\right) \cdot \mathbf{B}_{hkl} = 0 \qquad \left(\frac{\mathbf{a}_3}{l} - \frac{\mathbf{a}_2}{k}\right) \cdot \mathbf{B}_{hkl} = 0$$

and hence $\mathbf{B}_{hkl}$ is perpendicular to the planes hkl. The spacing of the planes hkl is

$$d_{hkl} = \left|\frac{\mathbf{a}_1}{h}\right| \cos \phi = \frac{\mathbf{a}_1}{h} \cdot \frac{\mathbf{B}_{hkl}}{|\mathbf{B}_{hkl}|} = \frac{1}{|\mathbf{B}_{hkl}|}$$

Equivalence of the three Laue equations and the Bragg law can be shown as follows: Any vector $\mathbf{r}$ can be expressed by

$$\mathbf{r} = (\mathbf{r} \cdot \mathbf{a}_1)\mathbf{b}_1 + (\mathbf{r} \cdot \mathbf{a}_2)\mathbf{b}_2 + (\mathbf{r} \cdot \mathbf{a}_3)\mathbf{b}_3 \tag{10}$$

Let $\mathbf{r}$ be the vector $(\mathbf{s} - \mathbf{s}_0)$ and combine it with the three Laue equations and Eq. (10) to obtain

$$\mathbf{s} - \mathbf{s}_0 = \lambda(\alpha\mathbf{b}_1 + \beta\mathbf{b}_2 + \gamma\mathbf{b}_3) \tag{11}$$

The Bragg law can be written in vector form as

$$\mathbf{s} - \mathbf{s}_0 = \lambda\mathbf{B}_{hkl} = \lambda(h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3) \tag{12}$$

since the usual form of the Bragg law is simply an equality in the magnitudes of the vectors: $|\mathbf{s} - \mathbf{s}_0| = 2 \sin \theta$ and $|\mathbf{B}_{hkl}| = 1/d_{hkl}$. Comparison of Eqs. (11) and (12) shows that the integers α, β, γ of the three Laue equations are simply the Miller indices hkl of the Bragg law.

The positions of the atoms in the unit cell are represented by a set of atomic coordinates x_n, y_n, z_n such that for atom n,

$$\mathbf{r}_n = \mathbf{a}_1 x_n + \mathbf{a}_2 y_n + \mathbf{a}_3 z_n \tag{13}$$

For a Bragg law reflection hkl, the structure factor takes the simple form

$$F_{hkl} = \sum_n f_n \exp\left[2\pi i(hx_n + ky_n + lz_n)\right] \tag{14}$$

Integrated intensity. In general, the intensity of a Bragg reflection, as expressed by Eq. (2), is not an experimentally measurable quantity. Other factors, such as the degree of mosaic structure in the crystal and the degree of parallelism of the primary beam, have a profound influence on the measured diffracted intensity for any setting of the crystal. To obtain measurements characteristic of the crystalline structure, it is necessary to adopt a more useful concept, the integrated intensity. For a small single crystal, one postulates that the crystal is to be turned at constant angular velocity ω through the Bragg law position, and that the total diffracted energy of the reflection is to be measured. The integrated intensity E is then given by

$$E = \iint I \frac{d\alpha}{\omega} \, dA \tag{15}$$

where $d\alpha$ is a change in orientation of the crystal and dA is an element of area at the point of observation.

Most of the equations used in x-ray diffraction studies are derived on the assumption that the intensity of the diffracted beam is so small that any

interaction with the primary beam can be neglected These are classed as the equations for the ideally imperfect crystal. For powder samples, in which the individual crystals are extremely small, [illegible]

For the ideally perfect crystal, it is necessary to use a more elaborate theory which allows for the interaction of *diffracted radiation with the primary* beam In general, it is the integrated intensity which is measured, and theory shows that the integrated intensity for an ideally imperfect crystal is larger than that for the ideally perfect crystal Many of the crystalline samples used for x-ray diffraction studies are not ideally imperfect, and the measured integrated intensity is accordingly less than that predicted by the ideally imperfect crystal formulas which are used in the interpretation. The situation is usually handled by *adding a correction factor called the extinction* correction to the formulas for the integrated intensity from the ideally imperfect crystal

Atomic coordinates. To have complete information about a crystalline structure, it is necessary to know all the atomic coordinates $x_n y_n z_n$ of the n atoms *comprising* the unit cell. The atomic coordinates *appear* in the structure factor as given by Eq (14), and sometimes the coordinates are obtained *directly* from structure factor values. Another way is to plot the electron density in the unit cell and infer the atomic *positions from* peaks in the electron density function. The electron density in the unit cell is given by the triple Fourier series

$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l F_{hkl}\exp\left[-2\pi i\left(\frac{hx}{a}+\frac{ky}{b}+\frac{lz}{c}\right)\right] \tag{16}$$

for which the coefficients are simply the structure factors F_{hkl} However, from experimental measurements of either an *intensity* or an integrated intensity, values for $|F_{hkl}|^2$ can be obtained These yield the magnitude of F_{hkl} but not the phase This is the most serious limitation to a straightforward determination of *crystalline* structures by x-ray diffraction methods The ambiguity in the phase of F_{hkl} prevents the *use* of the Fourier plot of Eq (16) as a general *method* for determining any crystalline structure

Simple structures are *uniquely* determined by combining the x ray intensity results with space group theory (see GROUP THEORY). The space group of a crystal is the repeating spatial arrangement of symmetry elements which the structure displays. Considering all the possible symmetry elements which can exist in a *crystalline* structure, group theory shows that there are only 230 essentially different possible combinations, and these constitute the 230 space groups. A knowledge of the macroscopic symmetry of the *crystal*, coupled with the systematically vanishing x-ray reflections, will usually determine the *space group*. The limitations imposed by the space group on the possible atomic positions, coupled with the limitations imposed by the measured $|F_{hkl}|^2$, often allow a complete and unique structure determination for not too complicated structures. For highly complex structures, it is never certain that x-ray diffraction analysis can yield a complete structure determination Additional techniques such as the isomorphous replacement by heavy atoms, the use of Patterson plots, and the determination of phase relations from inequalities are used with success on some of the complex structures.

Many structures of interest in solid-state physics exhibit various kinds of randomness and imperfections The precise nature of these is sometimes of more interest than the ideal average structure. Randomness and imperfections in a structure show themselves by producing a diffuse intensity in addition to the sharp Bragg reflections The temperature vibration of the atoms produces a diffuse *intensity called temperature diffuse scattering.* Quantitative measurements of this scattering lead to values for the velocity of high-frequency elastic waves and to a complete experimental determination of the spectrum of the elastic waves which constitute the thermal vibrations of the crystal. In alloys showing order-disorder changes, the short-range order parameters are obtained from quantitative measurement of the diffuse intensity which results from randomness in the atomic arrangement

CRYSTALLINE DIFFRACTION

Laue method. The Laue pattern uses polychromatic x-rays provided by the continuous spectrum from an x-ray tube operated at 35–50 kv. The transmission Laue pattern is obtained by passing a finely collimated beam through a thin single crystal and recording the diffracted beams on a photographic film placed several centimeters beyond the crystal. For each set of planes *hkl*, θ is *fixed and the Bragg law is satisfied by selecting* the proper λ from the primary beam. In a Laue pattern. the different diffracted beams have different wavelengths, and their directions are determined solely by the orientations of the *hkl* planes Transmission Laue patterns were once used for structure determinations, but their many disadvantages have made them practically obsolete.

On the other hand, the back reflection Laue pattern is used a great deal in the study of the orientation of crystals The back reflection Laue camera is shown schematically in Fig. 3 The polychromatic beam enters through a hole in the x-ray film and falls on a single crystal whose orientation can be set as desired by a system of goniometer circles Diffracted beams bent through angles 2θ approaching 180° are registered on the photographic film *For cubic crystals it is very easy to* read the crystal orientation from a back reflection Laue pattern, and the patterns find considerable use in the cutting of single-crystal metal ingots.

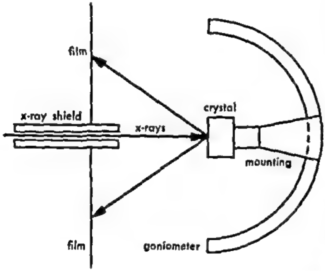


Fig. 3 Schematic representation of the back reflection Laue camera.

Rotating crystal method. The original rotating crystal method was employed in the Bragg spectrometer. A sufficiently monochromatic beam, of wavelength of the order of 1 A, is obtained by using the strong $K\alpha_1\alpha_2$ doublet with a filter which suppresses the $K\beta$ line and much of the continuous spectrum. The beam is collimated by a system of slits and then falls on the large extended face of a single crystal as shown by Fig. 4. Originally the diffracted beam was measured with an ionization chamber, but Geiger counters and proportional counters have largely replaced the ionization chamber. Both the crystal and the chamber turn about the spectrometer axis.

The Bragg spectrometer has been used extensively in obtaining quantitative measurements of the integrated intensity from planes parallel to the face of the crystal. The chamber is set at the correct 2θ angle with a slit so wide that all of the radiation reflected from the crystal can enter and be measured. The crystal is turned at constant angular speed ω through the Bragg law position, and the total diffracted energy E received by the ionization chamber during this process is measured. Similar readings with the chamber set on either side of the peak give a background correction. Fo this type of measurement, the integrated intensit E is given by

$$E = \frac{P_0}{2\mu\omega}\frac{e^4}{m^2c^4}\frac{\lambda^3F^2}{v^2}\left(\frac{1+\cos^2 2\theta}{2\sin 2\theta}\right)\exp[-2M] \quad (17$$

Where P_0 is the power of the primary beam, μ i the linear absorption coefficient in the crystal, ω i the angular velocity of the crystal, λ is the x-ra wavelength, F is the structure factor, v is th volume of the unit cell, and $\exp[-2M]$ is the so called Debye factor allowing for temperature vibra tion. When more than one kind of atom is presen this factor must be incorporated in F^2.

Measurements of the integrated intensity E giv quantitative values of F^2 directly. When a Geige counter is used in place of an ionization chambe for this type of measurement, it is necessary to employ a narrow counter-slit and traverse th counter through the reflected beam, since the sensi tivity of a Geiger counter is not constant over a large window opening.

The rotation camera, which is frequently use for structure determinations, is illustrated in Fig 5 The monochromatic primary beam s_0 falls on a small single crystal at O. The crystal is mounted with one of its axes (say a_3) vertical, and it rotates with constant velocity about the vertical axis during the exposure. The various diffracted beams are registered on a cylindrical film concentric with the axis of rotation For a rotation about a_3 it follows that $s_0 \cdot a_3 = 0$ and the third Laue equation gives

$$\sin\beta = \frac{l\lambda}{|a_3|} \quad (18)$$

The diffracted beams form the elements of a set of cones, and the intersection of these cones with the cylindrical film gives a set of horizontal lines of diffraction spots This type of pattern is called a rotation pattern, and the horizontal rows of spots are called layer lines. As seen from Eq. (18), the measured values of $\sin\beta$ give directly the length of the axis about which the crystal was rotated, and the layer line in which a spot occurs gives the l index of the reflection Similar rotations about the other two axes give corresponding information More elaborate variations of the rotation method, such as those of the Weissenberg and the precession cameras, involve a motion of the film in addition to the rotation of the crystal.

Powder method. The powder method involves the diffraction of a collimated monochromatic beam from a sample containing an enormous number of tiny crystals having random orientation Since about 1950, an increasing fraction of powder pattern studies have been made with Geiger counter, or proportional counter, diffractometers. The apparatus is shown schematically in Fig. 6 X-rays diverging from a target at T fall on the

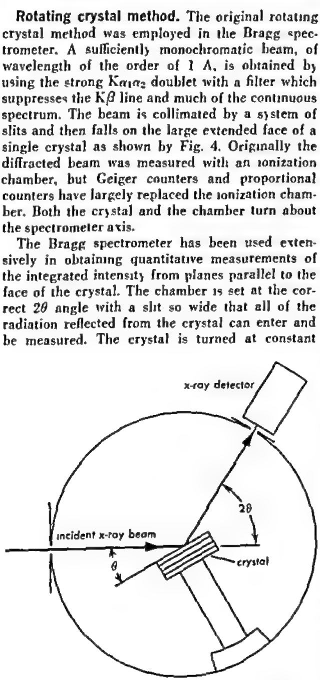


Fig. 4 Schematic representation of the Bragg spectrometer.

sample at O, the sample being a flat-faced briquet of powder Diffracted radiation from the sample passes through the receiving slit at s and enters the Geiger counter During the operation the sample turns at angular velocity ω and the counter at 2ω. The distances TO and OS are made equal to satisfy approximate focusing conditions A filter F before the receiving slit gives the effect of a sufficiently monochromatic beam. A chart recording of the amplified output of the Geiger counter gives directly a plot of intensity versus scattering angle 2θ For an extended discussion of this technique, see X-RAY POWDER METHODS

NONCRYSTALLINE DIFFRACTION

For a noncrystalline substance such as a glass or a liquid, a more general expression for the intensity of diffracted radiation is required. If the instantaneous position of each atom in the sample is represented by a vector $\mathbf{r}_n$, the diffracted intensity is given by

$$I = I_e \sum_q \sum_n f_q f_n \exp\left[\frac{2\pi}{\lambda}(\mathbf{s}-\mathbf{s}_0)\cdot(\mathbf{r}_q - \mathbf{r}_n)\right] \quad (19)$$

A particularly useful variation of Eq (19) is obtained by computing the average intensity $\langle I \rangle$ when the sample as a rigid array is allowed to take with equal probability all orientations in space.

$$\langle I \rangle = I_e \sum_q \sum_n f_q f_n \frac{\sin(kr_{qn})}{kr_{qn}} \quad (20)$$

where $k = 4\pi(\sin\theta)/\lambda$ and $r_{qn} = |\mathbf{r}_q - \mathbf{r}_n|$.

The fact that there are fairly definite nearest-neighbor and second neighbor distances in a glass or liquid means that Eq (20) will show peaks and dips when the intensity is plotted against $(\sin\theta)/\lambda$. Peaks and dips in an x-ray diffraction pattern merely indicate the existence of preferred interatomic distances, not that the material is necessarily

[illegible]

...ouring atoms at any distance from an average atom.

[B E W.]

Gases. Gases and liquids are found to give rise to x-ray diffraction patterns characterized by one or more halos or interference rings which are usually somewhat diffuse. These diffraction patterns, which are similar to those for glasses and amorphous solids, are due to interference effects depending both upon the electronic distribution of each of the individual atoms or molecules and upon their relative positions in the system.

For monatomic gases the only appreciable interference effects giving rise to a distribution of scattered intensities are those produced by the electronic distribution about each nucleus. These interference effects giving rise to so-called coherent intensities are the result of the interference of the individual waves scattered by electrons in different parts of the atom. The electronic distribution of an atom is described in terms of a characteristic atomic-scattering factor which is defined as the ratio of the resultant amplitude scattered by an atom to the amplitude that a free electron would scatter under the same conditions. At zero-angle scattering the atomic-scattering factor is equal to the atomic number of the atom. The coherent intensity in a given direction is proportional to the square of the atomic-scattering factor If it is assumed that the electronic distribution is spherically symmetrical, the atomic-scattering factors can be readily obtained from the observed intensities. For molecular gases the interference effects depend not only on the scattering factors of the atoms but also on their relative positions in the molecule. One can observe only an average intensity scattered over a period of time during which the molecules have taken innumerable positions with respect to the incident beam Interference effects due to the relative packing of the atoms or molecules can be neglected for dilute gases but not for dense gases

As in the case of x-ray diffraction by crystals, light atoms such as hydrogen are difficult to detect

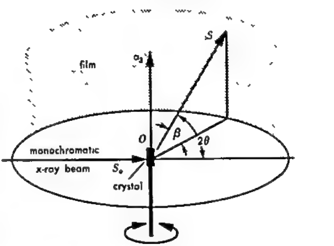


Fig 5 Schematic representation of the rotation camera

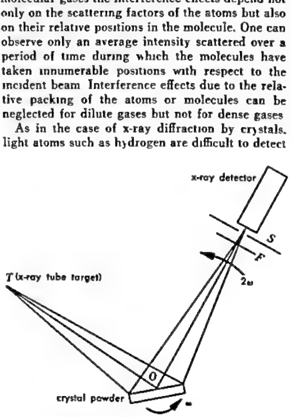


Fig 6 Schematic representation of the Geiger counter diffractometer for powder samples

in the presence of heavy atoms. Because of the shorter exposure times required, electron diffraction rather than x-ray diffraction has been utilized in studies of the structures of gaseous molecules. Both methods appear to be comparable in view of the accuracy of the intensity measurements and the technical difficulties involved.

Liquids. One cannot, as in the cases of dilute gases and crystalline solids, derive unambiguous, detailed descriptions of liquid structures from diffraction data. Nevertheless, diffraction studies of liquids do provide most useful information. Instead of comparing the experimental intensity distributions with theoretical distributions computed for various models, the experimental results are usually provided in the form of a radial distribution function which specifies the density of atoms or electrons as a function of the radial distance from any reference atom or electron in the system without any prior assumptions about the structure. From the radial distribution function one can obtain (1) the average interatomic distances most frequently occurring in the structure corresponding to the positions of the first, second, and possibly third nearest neighbors; (2) the distribution of distances; and (3) the average coordination number for each interatomic distance. The interpretation of these diffraction patterns given by the radial distribution function usually is not straightforward and in general one can say only that a certain assumed structural model and arrangement is not inconsistent with the observed diffraction data The models considered represent only a description of the time-average environment about any given atom or molecule within the liquid

There are great experimental difficulties in obtaining accurate intensity data. The sources of error are many; for a detailed treatment the reader is referred to publications of C Finbak. A brief description of some results obtained by x-ray diffraction of liquids is given below.

Liquid elements The radial distribution function, first used in a study of liquid mercury, has been applied to a considerable number of liquid elements mainly to compare their physical properties in the liquid and crystalline states. In most cases a lower first mean coordination number is found in the liquid state; exceptions are liquid gallium, bismuth, germanium, and lithium The radial distribution curves give direct evidence for the existence of molecules in some liquid elements (for example, N_2, O_2, Cl_2, and P_4) and imply the existence of more complicated atomic aggregates in a few cases. Argon and helium have been extensively studied in the liquid and vapor states over wide ranges of temperature and pressure.

Solutions. Until the 1950s x-ray diffraction had been applied only in a very qualitative way to the examination of structural arrangements of electrolytes in aqueous solutions. X-ray diffraction studies that were made in 1938 indicated that water has a broken-down ice structure in which each molecule is tetrahedrally bonded on an average to fewer than four nearest neighbors. Quantitative data on the average hydration numbers and on the average ion hydration distances are now being obtained. For detailed descriptions of the molecular structure of water, see HYDROGEN BOND; WATER. Radial distribution curves for concentrated $FeCl_3$ solutions indicate a large degree of local ordering of the ions with formation of Fe^{3+}-Cl^- complexes. Studies on metal-metal solutions, colloidal solutions, and molecular solutions have been made. More definite results have been obtained for concentrated solutions of strongly scattering solute in weakly scattering solvents. As examples may be cited the proof of the existence of a polymeric species in aqueous $Bi(ClO_4)_3$, evidence that in aqueous solution the HgX_4^{2-} anions ($X = Cl$, Br and I) are tetrahedral, and definite evidence of ion pair formation in aqueous BaI_2.

Molten salts. It is currently felt that a molten salt is a loose and expanded imitation of the solid with the same coordination scheme and short-range order Careful x-ray diffraction studies of a number of molten salts have indicated that these melts do not possess such quasi-crystalline structures but instead have quite open structures with a wide variety of individual ion coordinations. Interpretations of radial distribution functions for several other molten salts have been made. Liquid $AlCl_3$ appears to consist mainly of Al_2Cl_6 molecules; liquid SnI_4 is composed of independent tetrahedral molecules. The results for other molten salts are not as conclusive See ELECTRON DIFFRACTION; NEUTRON DIFFRACTION. [L F D.]

Bibliography · C S Barrett, *Structure of Metals*, 2d ed., 1952; W L. Bragg (ed), *The Crystalline State*, vol. 2, 1949, G. L. Clark, *Applied X-Rays*, 4th ed., 1955; B D Cullity, *Elements of X-Ray Diffraction*, 1956, C. Finbak, Structures of liquids, I–II, *Acta Chem Scand*, 3:1279–1308, 1949; D. McLachlan, *X-Ray Crystal Structure*, 1957; M. H Pirenne, *The Diffraction of X-Rays and Electrons by Free Molecules*, 1946; W. T. Sproull, *X-Rays in Practice*, 1946.

X-ray fluorescence analysis

A nondestructive physical method used for chemical analyses of solids and liquids. The specimen is irradiated by an intense x-ray beam which causes the elements in the specimen to emit (that is, fluoresce) their characteristic x-ray line spectra (see FLUORESCENCE) The lines of the spectra are diffracted at various angles by a single crystal plate which is analogous to the diffraction grating of optical spectroscopy The elements may be identified by the wavelengths of their spectral lines, which vary in a regular manner with atomic number, and their concentrations may be determined from the intensities of the lines. Counter tubes and associated electronic circuits are used to measure the x-ray intensities.

Unlike x-ray diffraction, the elements (rather than the compounds) are identified, and it is not necessary that the specimen be crystalline (see

X-RAY DIFFRACTION). The method supplements optical spectroscopy, being most generally used for concentrations $> 0.1\%$, although with prior wet chemical or physical extraction methods to enhance the peak-to-background ratio, much lower amounts can be detected

Analyses for all elements above about atomic number 12 can be done in a small fraction of the time required by conventional wet chemical methods. Aside from the initial cost of the equipment, x-ray fluorescence spectroscopy, which is also termed x-ray spectrochemical analysis, is inexpensive, and highly trained personnel are not usually required.

Although the method had its origin in the classic work of H G J Moseley in 1913, it was not widely used until about 1950, when the introduction of modern x-ray equipment made it feasible to apply the method to a large variety of analyses. The equipment and methods are quite similar to those employed in x-ray powder diffraction (*see* X-RAY POWDER METHODS).

Applications. The x-ray fluorescence method is less sensitive than the optical method, although with prior specimen treatment, analyses can be made in the range of parts per million. Microgram quantities have been determined in milligram samples with standard deviations of 10–15%, using focusing crystal spectrographs. Optical spectrographic methods become difficult to use for concentrations exceeding about 20%, whereas the x-ray method is applicable up to 100%.

The results agree well with those obtained by conventional wet chemical methods, while requiring only a small fraction of the time. For this reason, x-ray fluorescence methods have replaced chemical procedures for routine analyses in many fields. In metallurgy, for example analyses of tungsten [illegible]

[illegible] and. The time factor makes possible analyses on a very large scale, such as are required in geochemical problems where the costs for conventional methods would be prohibitive. Many applications to production control processes where speed is essential are now feasible

The x-ray fluorescence method has proven particularly useful for mixtures of elements of similar chemical properties which are difficult to separate and analyze by conventional chemical methods. The method has been applied to measuring the thickness of thin [illegible] evapora[illegible] by ana[illegible] operativ[illegible] applications of the method to liquids have been as numerous as to solids

Basis of the method. The origin of x-ray spectra may be understood from the simple Bohr model of the atom in which the electrons are arranged in orbits within the $K, L, M, \ldots$ shells (*see* ATOMIC STRUCTURE AND SPECTRA). If enough energy is applied to the atom, an electron may be ejected from

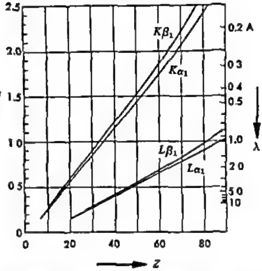


Fig 1. Plot of Moseley's law, showing dependence of characteristic x-ray line wavelengths on atomic number. (*Philips Tech. Rev* 17(10), 1956)

one of the inner shells An electron from one of the outer orbits promptly falls back to the inner shell to take its place, so that the atom returns to its normal state This action results in the emission of a characteristic x-ray spectral line, that is, a quantum of energy equivalent to the difference of the binding energies of the orbits involved in the electron transition Because only a limited number of electron orbits may participate, the x-ray spectrum of an element consists of comparatively few lines, in contrast to the much more complex optical spectra

Moseley's law Moseley showed that there was a linear relationship between $1/\sqrt{\lambda}$ and Z, where λ is the x-ray wavelength and Z is the atomic number of the element Figure 1 is a plot of Moseley's law for some K and L x-ray lines The K lines originate from electron transfers to the K shell and the L lines from transfers to the L shell The K lines have the shorter wavelengths and therefore the higher energies There are four or five lines in the K spectrum ($K\alpha_1$, $K\alpha_2$, $K\beta_1$, and so forth) and a dozen or more lines in the L spectrum ($L\alpha_1$, $L\alpha_2$, $L\beta_1$, and so forth) There are also M spectra, but these are of very long wavelengths and are rarely used in x-ray analysis In the filling of vacancies in the electron shells the ejection of Auger electrons is an alternative process to x-ray emission *See* AUGER EFFECT

Production of spectra A necessary condition for the production of a particular spectrum is that an energy called the critical excitation potential be exceeded This is the energy required to remove an electron from a particular shell of a given atom Thus all the L lines of copper appear above 1.1 thousand electron volts (kev), the K lines of copper above 8.98 kev, and the K lines of silver above 25.5 kev

There are two general methods for the produc-

tion of characteristic line spectra. In the first, the direct electron excitation method, primary excitation is used in the conventional x-ray tube, where electrons are focused onto a small portion, called the focal spot, of a pure target element. The generated line spectra are characteristic of the target material, and their intensities increase by nearly the square of the voltage (above the critical potential) and linearly with current. In addition to the line spectrum, there is also generated a continuous spectrum, resulting from the deceleration of the electrons (*see* BREMSSTRAHLUNG). The intensity and short wavelength limit of the continuous spectrum depend primarily on the applied voltage, as shown in Fig. 2.

The second method for the production of the characteristic line spectrum is the secondary excitation, or fluorescence method, so-called from analogy to the optical case. The specimen is placed outside a sealed-off x-ray tube and irradiated by the x-ray beam. All of the line and continuum primary x-rays having energies exceeding the critical excitation potential of the elements in the specimen may cause those elements to emit their characteristic x-ray line spectra. The usable fluorescence spectrum is only about 1/1000 as intense as the direct-electron-excited spectrum (the limitations are mainly due to the geometry of the spectrograph), but the high intensities and great sensitivities of modern equipment, and the convenience and simplicity of the method make it much more practical. Although there is no significant amount of continuous spectrum generated in x-ray fluorescence, there is always a small amount of scattered x-ray background.

X-ray absorption. Each element has an abrupt change in its x-ray absorption (called the absorp-

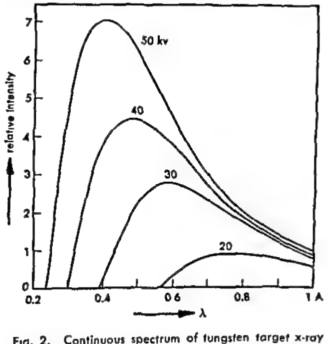


Fig. 2. Continuous spectrum of tungsten target x-ray tube obtained at various peak voltages (full-wave rectification) and same tube current. Measured with silicon 111 crystal analyzer and scintillation counter. (*Philips Tech. Rev.* 17(10), 1956)

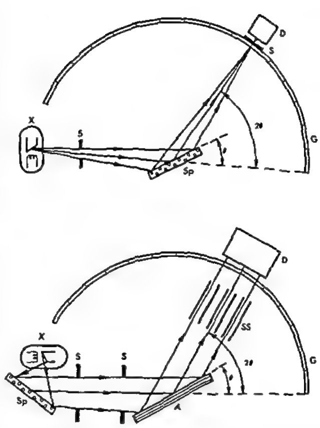


Fig. 3. Schematic drawing of x-ray powder diffractometer (above) and nonfocusing x-ray fluorescence spectrograph (below). *X*, x-ray tube; *S*, limiting slits; Sp, specimen, θ, glancing angle of incidence; 2θ, reflection angle, A, crystal analyzer; SS, Soller (parallel) slits; D, counter tube detector, G, goniometer.

tion edge) at the wavelength corresponding to the critical excitation potential for a given orbit. An element strongly absorbs x-rays of higher energy, *that is*, shorter wavelength, than its absorption edge. The amount of absorption is a function of wavelength; absorption decreases gradually below the absorption edge and increases again above the absorption edge. Thus the x-rays of wavelength just shorter than the absorption edge are most efficient in generating x-ray fluorescence, the efficiency decreasing as the exciting wavelength is further removed from the absorption edge. The wavelengths longer than the absorption edge have no effect in exciting fluorescence because they possess insufficient energy. In addition, the fluorescence conversion efficiency falls off with decreasing atomic number, because of, among other things, the Auger effect.

The absorption of x-rays increases rapidly with increasing wavelength, and because the line spectra of the low-atomic number elements are of comparatively long wavelength, these elements can be analyzed only with great difficulty or not at all. The attenuation of x-rays by the x-ray tube window and the air path through which the x-rays pass may be considerable, depending on the wavelengths involved. For example, an air path of 34 cm, which is

helium through

ments above about atomic number 24 (Cr) are usually done in air, while lower atomic numbers down to 12 (Mg) are done in a vacuum or a helium path Fluorescence analysis of elements below atomic number 12 is not possible with existing equipment.

Crystal analyzers. A single crystal plate is used to separate the various wavelengths emitted by the specimen. Diffraction from the crystal occurs according to Bragg's law:

$$n\lambda = 2d \sin \theta$$

where n is a small integer giving the order of reflection, λ the wavelength, d the spacing of the particular set of lattice planes of the crystal that are properly oriented to reflect, and θ the angle between those lattice planes and the incident ray. *See* X-RAY CRYSTALLOGRAPHY.

Reflection occurs only at an angle 2θ with respect to the incident ray, and it is therefore necessary to maintain the correct angular relationship of the crystal planes at one half the counter tube angle. The goniometer rotates the crystal at *one-half* the angular speed of the counter tube, and therefore both are always in the correct position to receive the various wavelengths emitted by the specimen as shown in Fig 3. For a given d there is only one angle (for each order of reflection) at which each wavelength is reflected, the angle increasing with wavelength

The angular separation of the lines, or the dispersion

$$d\theta/d\lambda = n/2d \sin \theta$$

increases with decreasing d. It is thus *easy to increase* the dispersion simply by selecting a crystal with a smaller d Reducing d also limits the maximum wavelength that can be measured *since* $\lambda = 2d$ at $2\theta = 180°$; the maximum 2θ *angle that* can be reached with the goniometer is about *150°*.

Although the fluorescent beam from the specimen is not parallel and has a rather large angular aperture, the crystal at any instant reflects only that bundle of parallel rays *making* the correct θ angle. As the crystal rotates, it *reflects* other bundles in sequence and *thereby* collimates the reflected beam. To increase the resolution, that is, decrease the line breadth, it is necessary to limit the angular range over which a wavelength is recorded. Parallel or Soller slits are used for this purpose as shown in Fig 3 These slits consist of thin (0.001-0.002 in) *equally* spaced foils of such materials as nickel and iron, *and* the angular aperture is determined by the length and spacing. A typical set would have 0.010-in spacing, 4-in. length with angular aperture 0.29°, and cross section about ¾ in square. The absorption of the foils is sufficiently high *to* prevent rays that are inclined by more than the angular aperture from entering the counter tube. Two sets of parallel slits

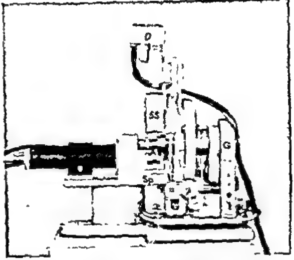


Fig. 4 Modern x-ray spectrograph for fluorescence analysis (Philips Electronics Co., Inc.)

may be used, one set between the specimen and crystal and the other between crystal and counter tube. This greatly increases the resolution and peak-to-background ratio, and causes a relatively small loss of peak intensity

It is essential that the crystal be of good quality to obtain sharp, symmetrical reflections Unless the crystal is homogeneous, the reflection may be distorted and portions of the reflections may occur at slightly different θ-angles Such effects would decrease the peak intensities of the different wavelengths by varying amounts, causing errors in the analysis

Instrumentation. The x-ray spectrograph, Fig. 4, for fluorescence analysis consists of the primary x-ray tube and its stabilized high-voltage power supply, a ray-proof specimen chamber; the goniometer which carries the crystal *analyzer and* counter tube, and the electronic circuitry *for the* counter tube (not shown) consisting *of a power* supply, linear amplifier, single-channel *pulse*-height analyzer, scaler, ratemeter, and *strip-chart* recorder.

X-ray tube operation The primary x-ray tube *targets are usually* tungsten, molybdenum, or copper It is usually necessary to *avoid the use of* a tube whose target is identical *with an element* in *the specimen* because the line spectrum from the *target is scattered* through the system It is also desirable *to select a* target whose characteristic lines *lie close to the absorption edges of the elements to be analyzed*; for example, the *WL* lines and the *CuK lines are* more efficient in exciting fluores*cence in the first transition* elements than are the MoK lines.

The focal spot size of the x-ray tube does not enter *directly* into the spectrograph geometry. Hence it is usually *made two or three* times larger *than those used* in diffraction tubes, and the tube *may thus* be operated at correspondingly higher *power* To obtain *the maximum* intensity, the specimen *should* be placed as close to the window of the tube as possible *since the intensity falls off* ac-

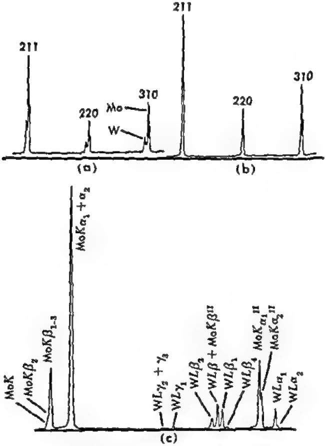


Fig 5. (a) Powder diffractometer diagram of Mo and W mixture. (b) Powder diffractometer diagram of Mo and W solid solution. (c) Fluorescence pattern of Mo and W. (*Svensk Kemisk Tidskrift 68, 1956*)

cording to the inverse square law In fluorescence analysis there is an important advantage in operating the x-ray tube at constant dc potential rather than full-wave rectification because x-rays of a certain wavelength are generated only during that fraction of the voltage-time cycle in which the critical excitation potential for that wavelength is exceeded. For example, the critical excitation potentials of the K spectra of Ni and Ba are 8 3 and 37.4 kev, respectively. If a primary x-ray tube with tungsten target is operated at 50 kilovolt peak full-wave, the fluorescent intensities of these elements will be only 0.60 and 0 27, respectively, of those obtained at 50 kv dc operation. *See* X-RAY TUBE

Equipment is normally operated at voltages up to 50–60 kv for fluorescence analysis These voltages can generate the K-spectra of all the elements up to the rare earths and the L-spectra of the higher atomic number elements, equipment operating up to 100 kv or more is required for the K-spectra of the latter group. Since counter tubes are used, it is essential to have a constant primary intensity and to stabilize the voltage and tube current.

Diffractometry The fluorescence-analysis instrumentation is similar to that used in x-ray powder diffractometry, as shown in Fig. 4. In diffractometry, x-rays of a single wavelength are used to measure the intensities corresponding to all the ds in the polycrystalline specimen. The diffraction pattern is characteristic of the structure of the element or compound. Figure 5*a* and *b* shows two x-ray diffractometer diagrams; Fig. 5*c* is an x-ray fluorescence pattern. Figure 5*a* is of a mixture of approximately 2 parts molybdenum and 1 part tungsten. Figure 5*b* shows a solid solution of Mo and W. Each reflection in the two diagrams consists of the $CuK\alpha_{1,2}$ doublet, a pair of closely spaced lines with a 2:1 intensity relationship. Figure 5*c* is an x-ray fluorescence pattern of Mo and W showing the MoK and WL spectra. The Roman numerals indicate second-order reflections.

Counter tubes Devices such as the Geiger-Müller counter and the scintillation counter are used almost exclusively rather than film as the x-ray detector because they make possible rapid and accurate intensity measurements In addition to such important characteristics as stability and reliability, counters for fluorescence analysis should have a linear response to avoid intensity corrections at the higher counting rates, a high quantum counting efficiency to obtain the maximum observable counting rate, and good energy resolution to allow discrimination against unwanted wavelengths. *See* GEIGER-MÜLLER COUNTER; SCINTILLATION COUNTER.

Intensity measurement. X-ray quanta are emitted in a random manner and hence certain statistical rules must be taken into account in using counter tubes for x-ray intensity measurements. There are three basic methods.

The fixed-time method uses the scaling circuit to measure the count N during a fixed predetermined time interval t and the counting rate n is determined from

$$n = N/t$$

This method is commonly used when the amount of time available is limited, but has the disadvantage of the error being dependent on the counting rate.

In the fixed-count method, N is predetermined and t is variable Although the error is then independent of n, long counting times will be required for low counting rates Both the fixed-time and fixed-count methods are essentially point-by-point manual procedures requiring the setting of the counter tube correctly on the peaks of the lines to be measured

A commonly used method is to operate the goniometer at a constant angular velocity and to record the output of the ratemeter with a strip chart recorder This introduces a distortion of the line profiles, causing the peaks to be lower than their correct intensity The decrease of peak intensity is proportional to the product of the time constant and the scanning speed of the goniometer. This method is employed for qualitative and semiquantitative analysis, and once the peaks are located, one of the manual methods may be used for a more precise measurement

Analytical methods. The specimens may be in the form of powders, briquettes, solids, or liquids The surface exposed to the primary x-ray beam

gree of surface roughness, which is difficult to measure quantitatively, causes losses in intensity and results in errors in the analysis. Consequently, solid samples are generally polished, and then if necessary are lightly etched or specially cleaned to remove contaminants. Powders of large particle size are generally ground and then briquetted, so that the surface will meet the conditions specified.

Normally, if the specimen is of unknown composition, a rate meter recording is used to make a qualitative or semiquantitative analysis. The reflection angles of each of the lines are read from the chart and the wavelengths computed using Bragg's equation or a table of 2θ vs λ for the specific crystal analyzer used. The elements are identified from the wavelengths, and the concentrations may be estimated from the relative intensities.

For accurate quantitative analysis, the usual procedure is to prepare a number of standard reference specimens of known composition for the various concentrations. The standards are used to prepare working calibration curves in which the intensities are plotted against the concentrations of each element in the required concentration range. The linearity and slope of the calibration curve are dependent upon the x-ray absorption characteristics of the specimen. The greater the differences in absorption of the elements to be analyzed, the closer together in composition should be the standards. In some analyses it is desirable to use an internal standard method, that is, to add a known amount of a reference element to the specimen during preparation. The calibration curve is then prepared by using for the intensity scale the ratio of the intensity of a line of the standard to that of another element of nearly the same absorption.

The intensities may be measured to the required statistical accuracy at the peak of the line using a fixed-count or fixed time method. If the lines have different shapes, as for example in comparing an unresolved with a partially resolved doublet, the integrated line intensities should be obtained by scanning across the lines and totalizing the counts with the scaler.

The precision of the method depends upon many factors: specimen preparation and homogeneity, the accuracy of the calibration curves, the stability of the x-ray source, counting statistics, and the resolution and dispersion of the spectrograph. In some cases overlapping lines may be difficult to eliminate and may influence the precision. With a good modern properly adjusted instrument, measurements of elements present in amounts more than 1% can usually be made in a few minutes to a precision of a few per cent or better. [W P]

Bibliography: H. Friedman and L. S. Birks, A Geiger counter for x-ray fluorescence analysis, *Rev. Sci. Instr.*, 19(5):323–330, 1948. W. Parrish, X-ray spectrochemical analysis, *Philips Tech. Rev.*, 17(10):269–286, 1956; N. Spielberg, W. Parrish and K. Lowitzsch, Geometry of the nonfocusing x-ray fluorescence spectrograph, *Spectrochim. Acta*, vol. 13 (August), 1959.

X-ray optics

A title-by-analogy of those phases of x-ray physics in which x-rays demonstrate properties similar to those of light waves. X-ray optics is also called roentgen optics.

Optics is that branch of physical science which deals with the nature and properties of light and the laws of its modification by opaque and transparent bodies. Essentially, then, it involves the electromagnetic wave phenomena of reflection from mirrors and of refraction or change in direction of a beam in passing the boundary between two media, as from air into a glass prism, thus spreading white light out into a rainbow or spectrum of colors. It also includes the diffraction or bending around opaque obstacles such as slits or the lines in ruled gratings, the spreading of light into a spectrum, and the polarization or constraint of vibrations in all directions [illegible]

wave [illegible]

crys[illegible]

tics [illegible]

ing of the paths of light rays, with image formation, and with instrumentation, such as mirrors, lenses in microscopes and telescopes, refractometers, polarimeters, interferometers, and spectrometers with prisms and ruled gratings. See OPTICS; OPTICS, GEOMETRICAL; OPTICS, PHYSICAL.

X-rays, when first discovered by W. K. Rontgen, seemed to possess none of the optical properties mentioned. See X-RAY(S), PHYSICAL NATURE OF. Later, however, polarization, diffraction, reflection, and refraction were all detected by using crystals having refractive indices that were slightly less than 1. Total reflection or diffraction of x-rays from ruled gratings could occur only with incident beams at very small grazing angles of the order of a few minutes. Until this discovery by A. Compton in 1922 all attempts to concentrate or focus x-rays by lenses and mirrors had failed.

X-ray microscopes. In 1948, accepting the fact of total reflection at very small grazing angles, P. Kirkpatrick devised the first true x-ray microscope (see MICROSCOPE, X-RAY). A concave spherical mirror receiving x-rays at grazing incidence images a point into a line in accordance with the focal length $f = Ri/2$, where R is the radius of curvature and i the grazing angle. The image is subject to an aberration such that a ray reflected at the periphery of the mirror misses the focal point [illegible]

[illegible] resolving power actually achieved by Kirkpatrick was such as to resolve points separated by 70 A, independent of wavelength.

Point images of points, and therefore extended images of extended objects, may be produced by causing x-rays to reflect from two concave mirrors

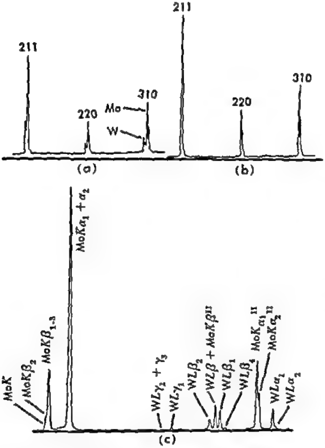


Fig 5. (a) Powder diffractometer diagram of Mo and W mixture (b) Powder diffractometer diagram of Mo and W solid solution (c) Fluorescence pattern of Mo and W. (*Svensk Kemisk Tidskrift 68, 1956*)

cording to the inverse square law In fluorescence analysis there is an important advantage in operating the x-ray tube at constant dc potential rather than full-wave rectification because x-rays of a certain wavelength are generated only during that fraction of the voltage-time cycle in which the critical excitation potential for that wavelength is exceeded. For example, the critical excitation potentials of the K spectra of Ni and Ba are 8 3 and 37.4 kev, respectively. If a primary x-ray tube with tungsten target is operated at 50 kilovolt peak full-wave, the fluorescent intensities of these elements will be only 0.60 and 0.27, respectively, of those obtained at 50 kv dc operation. *See* X-RAY TUBE

Equipment is normally operated at voltages up to 50–60 kv for fluorescence analysis. These voltages can generate the K-spectra of all the elements up to the rare earths and the L-spectra of the higher atomic number elements; equipment operating up to 100 kv or more is required for the K-spectra of the latter group Since counter tubes are used, it is essential to have a constant primary intensity and to stabilize the voltage and tube current.

Diffractometry The fluorescence-analysis instrumentation is similar to that used in x-ray powder diffractometry, as shown in Fig. 4 In diffractometry, x-rays of a single wavelength are used to measure the intensities corresponding to all the ds in the polycrystalline specimen. The diffraction pattern is characteristic of the structure of the element or compound. Figure 5a and b shows two x-ray diffractometer diagrams; Fig. 5c is an x-ray fluorescence pattern. Figure 5a is of a mixture of approximately 2 parts molybdenum and 1 part tungsten. Figure 5b shows a solid solution of Mo and W. Each reflection in the two diagrams consists of the $CuK\alpha_{1,2}$ doublet, a pair of closely spaced lines with a 2:1 intensity relationship. Figure 5c is an x-ray fluorescence pattern of Mo and W showing the MoK and WL spectra. The Roman numerals indicate second-order reflections.

Counter tubes. Devices such as the Geiger-Muller counter and the scintillation counter are used almost exclusively rather than film as the x-ray detector because they make possible rapid and accurate intensity measurements In addition to such important characteristics as stability and reliability, counters for fluorescence analysis should have a linear response to avoid intensity corrections at the higher counting rates, a high quantum counting efficiency to obtain the maximum observable counting rate, and good energy resolution to allow discrimination against unwanted wavelengths. *See* GEIGER-MULLER COUNTER; SCINTILLATION COUNTER

Intensity measurement. X-ray quanta are emitted in a random manner and hence certain statistical rules must be taken into account in using counter tubes for x-ray intensity measurements There are three basic methods

The fixed-time method uses the scaling circuit to measure the count N during a fixed predetermined time interval t and the counting rate n is determined from

$$n = N/t$$

This method is commonly used when the amount of time available is limited but has the disadvantage of the error being dependent on the counting rate

In the fixed-count method, N is predetermined and t is variable Although the error is then independent of n, long counting times will be required for low counting rates Both the fixed-time and fixed-count methods are essentially point-by-point manual procedures requiring the setting of the counter tube correctly on the peaks of the lines to be measured.

A commonly used method is to operate the goniometer at a constant angular velocity and to record the output of the ratemeter with a strip chart recorder This introduces a distortion of the line profiles, causing the peaks to be lower than their correct intensity The decrease of peak intensity is proportional to the product of the time constant and the scanning speed of the goniometer This method is employed for qualitative and semiquantitative analysis, and once the peaks are located, one of the manual methods may be used for a more precise measurement

Analytical methods. The specimens may be in the form of powders, briquettes, solids, or liquids The surface exposed to the primary x-ray beam

must be flat, smooth, and representative of the sample as a whole because usually only a thin surface layer contributes to the fluorescent beam. The degree of surface roughness, which is difficult to measure quantitatively, causes losses in intensity and results in errors in the analysis. Consequently, solid samples are generally polished, and then if necessary are lightly etched or specially cleaned to remove contaminants. Powders of large particle size are generally ground and then briquetted, so that the surface will meet the conditions specified.

Normally, if the specimen is of unknown composition, a rate-meter recording is used to make a qualitative or semiquantitative analysis. The reflection angles of each of the lines are read from the chart and the wavelengths computed using Bragg's equation or a table of 2θ vs λ for the specific crystal analyzer used. The elements are identified from the wavelengths, and the concentrations may be estimated from the relative intensities.

For accurate quantitative analysis, the usual procedure is to prepare a number of standard reference specimens of known composition for the various concentrations. The standards are used to prepare working calibration curves in which the intensities are plotted against the concentrations of each element in the required concentration range. The linearity and slope of the calibration curve are dependent upon the x-ray absorption characteristics of the specimen. The greater the differences in absorption of the elements to be analyzed, the closer together in composition should be the standards. In some analyses it is desirable to use an internal standard method, that is, to add a known amount of a reference element to the specimen during preparation. The calibration curve is then prepared by using for the intensity scale the ratio of the intensity of a line of the standard to that of another element of nearly the same absorption.

The intensities may be measured to the required statistical accuracy at the peak of the line using a fixed-count or fixed time method. If the lines have different shapes, as for example in comparing an unresolved with a partially resolved doublet, the integrated line intensities should be obtained by scanning across the lines and totalizing the counts with the scaler.

The precision of the method depends upon many factors: specimen preparation and homogeneity, the accuracy of the calibration curves, the stability of the x-ray source, counting statistics, and the resolution and dispersion of the spectrograph. In some cases, overlapping lines may be difficult to eliminate and may influence the precision. With a good modern, properly adjusted instrument, measurements of elements present in amounts more than 1% can usually be made in a few minutes to a precision of a few per cent or better. [W.P.]

Bibliography: H. Friedman and L. S. Birks, A Geiger counter for x-ray fluorescence analysis, *Rev. Sci. Instr.*, 19(5):323-330, 1948; W. Parrish, X-ray spectrochemical analysis, *Philips Tech. Rev.*, 17(10):269-286, 1956; N. Spielberg, W. Parrish, and K. Lowitzsch, Geometry of the nonfocusing x-ray fluorescence spectrograph, *Spectrochim. Acta*, vol. 13 (August), 1959.

X-ray optics

A title-by-analogy of those phases of x-ray physics in which x-rays demonstrate properties similar to those of light waves. X-ray optics is also called roentgen optics.

Optics is that branch of physical science which deals with the nature and properties of light and the laws of its modification by opaque and transparent bodies. Essentially, then, it involves the electromagnetic wave phenomena of reflection from mirrors and of refraction or change in direction of a beam in passing the boundary between two media, as from air into a glass prism, thus spreading white light out into a rainbow or spectrum of colors. It also includes the diffraction or bending around opaque obstacles such as slits or the lines in ruled gratings, the spreading of light into a spectrum, and the polarization or constraint of vibrations in all directions transverse to the direction of wave propagation into one direction, as in certain crystals, or by grazing reflection from surfaces. Optics thus is concerned with the control and focusing of the paths of light rays, with image formation, and with instrumentation, such as mirrors, lenses in microscopes and telescopes, refractometers, polarimeters, interferometers, and spectrometers with prisms and ruled gratings. See OPTICS; OPTICS, GEOMETRICAL; OPTICS, PHYSICAL.

X-rays, when first discovered by W. K. Röntgen, seemed to possess none of the optical properties mentioned. See X-RAY(S), PHYSICAL NATURE OF. Later, however, polarization, diffraction, reflection, and refraction were all detected by using crystals having refractive indices that were slightly less than 1. Total reflection or diffraction of x-rays from ruled gratings could occur only with incident beams at very small grazing angles of the order of a few minutes. Until this discovery by A. Compton in 1922 all attempts to concentrate or focus x-rays by lenses and mirrors had failed.

X-ray microscopes. In 1948, accepting the fact of total reflection at very small grazing angles, P. Kirkpatrick devised the first true x-ray microscope (see MICROSCOPE, X-RAY). A concave spherical mirror receiving x-rays at grazing incidence images a point into a line in accordance with the focal length $f = Ri/2$, where R is the radius of curvature and i the grazing angle. The image is subject to an aberration such that a ray reflected at the periphery of the mirror misses the focal point of central rays by a distance $S = 1.5 Mr^2/R$, where M is the magnification of the mirror and r the radius of the mirror face. The possible resolving power actually achieved by Kirkpatrick was such as to resolve points separated by 70 A, independent of wavelength.

Point images of points, and therefore extended images of extended objects, may be produced by causing x-rays to reflect from two concave mirrors

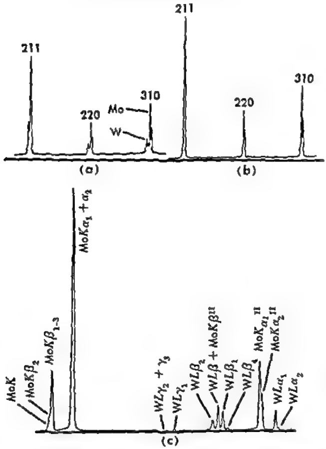


Fig. 5 (a) Powder diffractometer diagram of Mo and W mixture. (b) Powder diffractometer diagram of Mo and W solid solution. (c) Fluorescence pattern of Mo and W. (*Svensk Kemisk Tidskrift 68, 1956*)

cording to the inverse square law In fluorescence analysis there is an important advantage in operating the x-ray tube at constant dc potential rather than full-wave rectification because x-rays of a certain wavelength are generated only during that fraction of the voltage-time cycle in which the critical excitation potential for that wavelength is exceeded For example, the critical excitation potentials of the K spectra of Ni and Ba are 8 3 and 37.4 kev, respectively. If a primary x-ray tube with tungsten target is operated at 50 kilovolt peak full-wave, the fluorescent intensities of these elements will be only 0 60 and 0.27, respectively, of those obtained at 50 kv dc operation. *See* X-RAY TUBE.

Equipment is normally operated at voltages up to 50–60 kv for fluorescence analysis. These voltages can generate the K-spectra of all the elements up to the rare earths and the L-spectra of the higher atomic number elements; equipment operating up to 100 kv or more is required for the K-spectra of the latter group Since counter tubes are used, it is essential to have a *constant primary* intensity and to stabilize the voltage and tube current

Diffractometry. The fluorescence-analysis instrumentation is similar to that used in x-ray powder diffractometry, as shown in Fig 4. In diffractometry, x-rays of a single wavelength are used to measure the intensities corresponding to all the *d*s in the polycrystalline specimen The diffraction pattern is characteristic of the structure of the element or compound. Figure 5*a* and *b* shows two x-ray diffractometer diagrams; Fig. 5*c* is an x-ray fluorescence pattern. Figure 5*a* is of a mixture of approximately 2 parts molybdenum and 1 part tungsten. Figure 5*b* shows a solid solution of Mo and W. Each reflection in the two diagrams consists of the $CuK\alpha_{1,2}$ doublet, a pair of closely spaced lines with a 2:1 intensity relationship. Figure 5*c* is an x-ray fluorescence pattern of Mo and W showing the MoK and WL spectra. The Roman numerals indicate second-order reflections.

Counter tubes. Devices such as the Geiger-Muller counter and the scintillation counter are used almost exclusively rather than film as the x-ray detector because they make possible rapid and accurate intensity measurements. In addition to such important characteristics as stability and reliability, counters for fluorescence analysis should have a linear response to avoid intensity corrections at the higher counting rates, a high quantum counting efficiency to obtain the maximum observable counting rate. and good energy resolution to allow discrimination against unwanted wavelengths. *See* GEIGER-MÜLLER COUNTER; SCINTILLATION COUNTER

Intensity measurement. X-ray quanta are emitted in a random manner and hence certain statistical rules must be taken into account in using counter tubes for x-ray intensity measurements. There are three basic methods.

The fixed-time method uses the scaling circuit to measure the count N during a fixed predetermined time interval t and the counting rate n is determined from

$$n = N/t$$

This method is commonly used when the amount of time available is limited, but has the disadvantage of the error being dependent on the counting rate.

In the fixed-count method, N is predetermined and t is variable. Although the error is then independent of n, long counting times will be required for low counting rates. Both the fixed-time and fixed-count methods are essentially point-by-point manual procedures requiring the setting of the counter tube correctly on the peaks of the lines to be measured.

A commonly used method is to operate the goniometer at a constant angular velocity and to record the output of the ratemeter with a strip chart recorder This introduces a distortion of the line profiles, causing the peaks to be lower than their correct intensity The decrease of peak intensity is proportional to the product of the time constant and the scanning speed of the goniometer This method is employed for qualitative and semiquantitative analysis, and once the peaks are located, one of the manual methods may be used for a more precise measurement.

Analytical methods. The specimens may be in the form of powders, briquettes, solids, or liquids The surface exposed to the primary x-ray beam

the diffraction data. The x-ray tube is sealed off (that is, has a permanent vacuum) and has a water-cooled target and three or four low-energy-absorbing mica or beryllium windows. X-ray cameras may be mounted at other windows.

X ray spectra. Only about six elements are useful for x ray tube targets. The spectrum of copper, the most commonly used target, is shown in Fig. 2a. It consists of a few lines superimposed on a broad continuum. These lines are the intense Cu$K\alpha$ doublet and the weaker Cu$K\beta$ triplet. In the doublet, the $K\alpha_1$ line at 1.541 A is about twice as intense as [illegible]

[illegible]

wavelengths of the lines are determined by the target element, and their intensity by the voltage and current applied to the x-ray tube. The other weak lines in this figure are caused by small amounts of tungsten impurity. There are also L and M spectral lines of much longer wavelengths, but these are absorbed by the x-ray tube window. The continuum begins at a wavelength which depends on the applied voltage, rises rapidly to a broad maximum, and then diminishes with increasing wavelength. The intensity of the continuum is mainly dependent on the x ray tube voltage, but it also increases with atomic number of the target element. *See* X RAY TUBE.

The entire spectrum is scattered and diffracted with varying efficiencies by the polycrystalline specimen. The interpretation of the diffraction pattern is simplified if the $K\beta$ line is eliminated. This is easily accomplished by inserting in the x ray path a thin nickel foil which almost completely absorbs the $K\beta$ radiation and transmits the $K\alpha$ as shown in Fig 2b. Other filter elements are used for other targets. To decrease the background from the continuum, and thereby increase the peak-to-background ratio of the diffraction pattern, a proportional or scintillation counter may be used in conjunction with a single-channel pulse-height analyzer. When the analyzer window is centered on the Cu$K\alpha$ lines and narrowed to pass about 90% of Cu$K\alpha$ radiation, the recorded spectrum consists almost entirely of Cu$K\alpha$, as shown in Fig. 2c.

Lattice geometry. The atoms in crystalline substances are arranged in a symmetrical three-dimensional pattern in which some atomic arrangement is repeated by the symmetry of the crystal along straight lines throughout the crystal. The array of points and the light lines in Fig 3 outline a lattice or framework of a typical crystal in which the third dimension is normal to the plane of the drawing. The smallest group of atoms which has the symmetry of the entire pattern is called the unit cell. There are several ways in which the unit cell is selected. In this case it was drawn parallel to the crystallographic axes a, b, and c. The traces of the various planes (normal to the drawing) are indicated by heavy lines. The method used by crystallographers to identify these planes is as follows: for a given set of planes count the number of planes crossed from one lattice point to the next

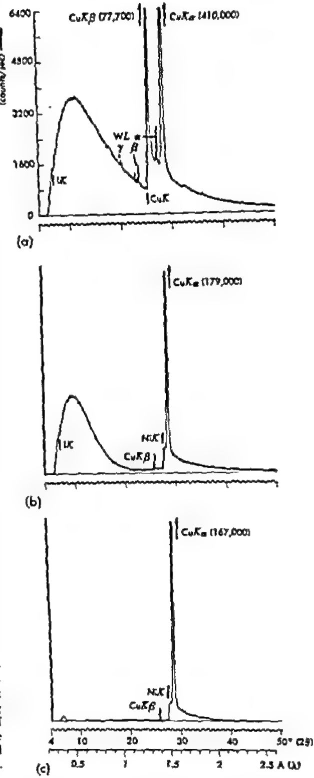


Fig 2 Spectrum of copper target x-ray tube operated at 40 kilovolts peak (kvp), full-wave rectification. Spectrum recorded using a silicon crystal plate cut parallel to the (111) planes and an NaI(Tl) scintillation counter. (a) Spectrum from tube (b) Same as (a) but with nickel filter 15 microns thick. (c) Same as (b) but with pulse-height analyzer set to pass about 90% of Cu$K\alpha$. (Modified from Int Tables for X-Ray Cryst, vol 3)

in series crossed at right angles to each other. Elliptical gold mirrors were found superior to spherical ones. This gave the impetus to a number of other varieties of x-ray microscopy, including contact and projection microradiography (*see* MICRORADIOGRAPHY).

An outstanding success of this reflecting optical system is in the isolation and focusing of very soft (low-energy) x-rays with wavelengths up to 45 A (the $K\alpha$ line of copper) in the remarkable work of B. Henke on microradiography of single blood cells.

Beam formation. Bent crystals can be used as diffraction gratings to focus diffracted beams and to form enlarged images of a specimen, itself serving as the source of x-rays or bathed by a beam of x-rays, in accordance with the same principles and equations that apply to light optics.

In x-ray fluorescent emission spectrometers, where it is necessary to concentrate beams of low intensities, bent reflecting or transmitting crystals (for example, mica) are so arranged that the theoretical curvature required can be varied with the diffraction angle of a spectrum line (*see* X-RAY FLUORESCENCE ANALYSIS). Focusing powder diffraction cameras, to gain maximum intensities, have the specimen and the photographic film on the same circumference, so that a larger area of sample can be irradiated with focusing of all diffracted wavelets into a sharp interference (*see* X-RAY DIFFRACTION; X-RAY POWDER METHODS). Optical principles are involved in assuring sharp-edged shadows in the projection x-ray microscope, which uses x-rays from a point source. Inherent enlargement without loss of detail is gained on images recorded on film at large distances from the specimen

Continuing effort is being made to improve the collimation of x-ray beams into finer pencils while retaining the maximum intensity of the radiation. Reflection from walls under grazing-angle conditions is such a method. Thus calcite-faced slits, very fine lead-glass capillary tubes, converging polished walls, bent optical flats, and similar devices have been used to produce beams smaller than 1 micron in width with 200 times the intensity of a slit system having the same definition at the same working distance, while at the same time these beams are partly monochromatized. Conversely, the total reflection of x-rays may be used to investigate the structure of optical surfaces from the widening of images; even the best polished surfaces consist of hills and valleys about 10 A in height and 1 μ in width. *See* COMPTON EFFECT; X-RAY CRYSTALLOGRAPHY. [G.L.CL.]

X-ray powder methods

Physical techniques used for the identification of substances and for other types of analyses, principally for crystalline materials in the solid state. In these techniques, a small collimated beam of nearly monochromatic x-rays is directed onto a small polycrystalline specimen in the form of powder, producing a diffraction pattern that is recorded on film or with a counter tube. This x-ray pattern is a fundamental and uniquely characteristic property resulting from the atomic arrangement of the diffracting substance. Different substances have different atomic arrangements or crystal structures, and hence no two chemically distinct substances give identical diffraction patterns. Identification may be made by comparing the pattern of the unknown substance with patterns of known substances, in a manner somewhat analogous to the identification of persons by their fingerprints. The analytical information is different from that obtained by chemical or spectrographic analysis. X-ray identification of chemical compounds indicates the constituent elements, and shows how they are combined.

The x-ray powder method is widely used in fundamental and applied research; for instance, it is used in the analysis of raw materials and finished products, in phase diagram investigations, in following the course of solid-state chemical reactions, and in the study of minerals, ores, rocks, metals, chemicals, and many other types of material. The use of powder methods to determine the actual atomic arrangement, which has been so important in the study of chemical bonds, crystal physics, and crystal chemistry, is described in related articles. *See* X-RAY CRYSTALLOGRAPHY, X-RAY DIFFRACTION; X-RAY FLUORESCENCE ANALYSIS.

Instrumentation. Complete automatic equipment for x-ray analysis is available from several companies A typical counter tube installation is shown in Fig 1 The high-voltage generator which provides rectified voltage for the x-ray tube is stabilized so that the x-ray intensity varies by less than 1% The diffractometer goniometer is mounted on the table in front of the x-ray tube window. The electronic circuits mounted in the rack on the right are used to operate the Geiger, proportional, or scintillation counter x-ray detector, and to record

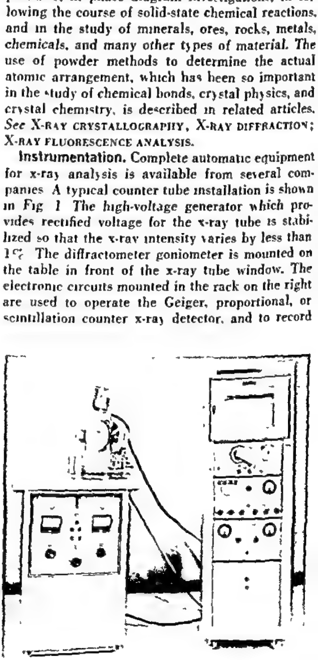


Fig 1 Modern x-ray diffraction equipment, showing high-voltage control (*left*) with x-ray tube housing and diffractometer goniometer on table, and recording circuits (*right*) (*Philips Electronics*)

not already exist in polycrystalline form, it may be pulverized When a fine-grained powder consisting of thousands of small, randomly oriented crystallites is exposed to the x ray beam, all the possible reflections from the various sets of atomic planes can occur simultaneously.

Recording of powder patterns. The two principal methods of recording diffraction patterns are by the use of film and with x-ray counter tubes. In the film method, a small amount of the powder is glued to a thin glass fiber and rotated continuously in the center of a powder camera. The latter is essentially a lighttight cylindrical enclosure, usually about 4.5 in. in diameter and provided with collimators to define the x-ray beam. The strip of x ray film is placed against the inside circumference of the cylinder concentric with the specimen axis. Two holes are punched in the film, one to admit the collimator, and the other to allow the undiffracted beam to pass through. Exposures of 1–4 hours are required Typical films are shown in Fig 4.

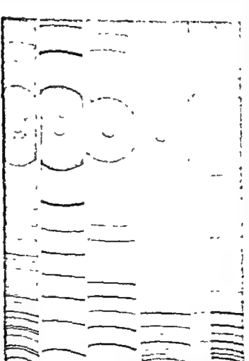


Fig 4 Films of several polycrystalline substances obtained with a cylindrical powder camera and $CuK\alpha$ x-rays. (a) $Pb(NO_3)_2$ (b) W (c) NaCl (d) SiO_2-α quartz. (e) Same as (d) but with Ni filter covering one side of film (*Science*, 110(2858) 349, 1949).

Each properly oriented crystallite may produce a reflection that appears as a spot on the film For each set of atomic planes there will be a large number of crystallites of various orientations, each producing its own spot, and all the reflected rays will form the surface of a cone, with its apex at the specimen and subtending the angle 4θ on the film The various sets of atomic planes thus produce a series of concentric cones and these appear on the film as a series of arcs whose curvature depends on the reflection angle. After the film is developed and dried it is laid flat and the linear distance measured from the position on the film where the direct beam passed through to each of the arcs. Since λ and the camera diameter are known and since allowance may be made for film shrinkage in development, these linear distances may be converted into angular values, from which *d* may be calculated for each set of planes by substituting in the Bragg equation

A much more powerful instrument, called the diffractometer (Fig 1), was developed in 1944; it employs a counter tube instead of film. The specimen is prepared with a flat surface which is exposed to the x-ray beam The counter tube always points toward the specimen, and the reflection data are recorded line by line The divergent primary x-ray beam converges, after reflection from the specimen, to a narrow slit in front of the counter tube By rotating the counter tube at twice the angular speed of the specimen so that the specimen surface always makes an angle θ with the primary beam when the counter tube is at 2θ sharp well-resolved lines are obtained at all reflection angles This instrument makes possible the direct and accurate measurement of intensities and reflection angles The peak-to-background ratios, resolution, and line shapes are far superior to those obtained by the cylindrical film camera just described The operation is completely automatic.

Identification of crystals. The most important use of the powder method is for the *identification* of crystalline substances Because no two different substances give identical powder patterns an identification may be made by comparison of the unknown pattern with knowns. In the United States, sets of 3- by 5-in cards listing the pertinent information, such as the *d* values and relative intensities of the lines of many thousands of substances, are published by the American Society for Testing Materials

The complexity of the pattern is determined primarily by the symmetry of the substance rather than by its chemical composition Hence, chemically complex compounds such as $(Cu_{10}Zn_2)Sb_4O_{13}$, $(NH_4)_3AlF_6$, and $2Na_2SO_4 \cdot NaCl \cdot NaF$ have patterns that are nearly as simple as those from Fe or Cu, because all are cubic. Thus the simplest patterns that is, those with the smallest number of lines, occur in the cubic system and the number of lines increases progressively, with decreasing symmetry becoming greatest in the triclinic system The number of lines also increases with unit cell

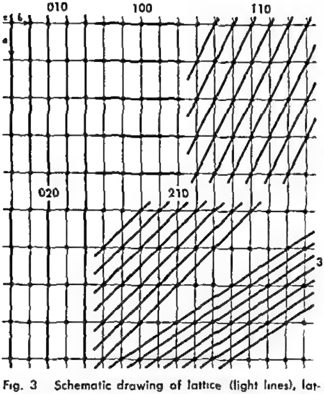


Fig. 3 Schematic drawing of lattice (light lines), lattice planes (heavy lines), and their Miller indices The orthogonal axes a and b outline the simplest unit cell, and the c axis is normal to the paper

along a, then repeat the procedure along b and c. The resulting numbers, called h, k, l, respectively, are known as the Miller indices of that set of planes, and the assignment of indices to each of the lines is called indexing the pattern. The Miller indices of the (102) planes are called one, zero, two.

The spacings d between the planes are related to the Miller indices and the unit cell dimensions. In crystals of the cubic system, the crystallographic axes are normal to each other and have the same length, $a = b = c$, and

$$d = a/(h^2 + k^2 + l^2)^{1/2}$$

Similar relations exist for the other five crystal systems The relations for the low-symmetry systems, monoclinic and triclinic, are much more complicated. Various types of charts to facilitate indexing cubic, tetragonal, and hexagonal substances are available. See CRYSTALLOGRAPHY; CRYSTAL STRUCTURE.

Bragg's law of reflection. *The atomic diameters* and the wavelengths of x-rays used have approximately the same dimensions, and therefore the crystal acts as a three-dimensional grating for x-rays in a manner analogous to the diffraction of ultraviolet or visible light by a ruled one-dimensional grating (see DIFFRACTION GRATING). Under appropriate conditions, the electrons around each atom scatter the incident x-ray beam in a coherent manner and in certain specific directions; the scattering from billions of atoms is in phase at the same time. Shortly after the discovery of x-ray diffraction by M. von Laue in 1912 this complex phenomenon was formulated in a simple relation by W. H. and W. L. Bragg, in which the diffraction is visualized as a reflection from a large number of parallel planes. (This should not be confused with the total reflection of x-rays, which occurs at very small grazing angles from highly polished surfaces.) They showed that when two or more parallel rays of the same wavelength are incident at the same glancing angle θ to a set of atomic planes, the path difference of the reflected rays from adjacent planes was 1 wavelength. This may be expressed as $\lambda = 2d_{hkl} \sin \theta$, where λ is the wavelength of the incident x-rays and d_{hkl} the interplanar spacing between the (hkl) planes of atoms. Thus the conditions for x-ray reflection are very restrictive because there is only one angle θ at which the x-rays of a given wavelength are reflected by a particular set of atomic planes of spacing d.

The reflection law may be illustrated by use of Cu$K\alpha$ x-rays and a flat single crystal section of sodium chloride cleaved parallel to the cube face, $(hkl) = (100)$ with $d = 5.64$ A. When this crystal surface is correctly oriented on the goniometer at $\theta = 7.8°$, and the counter tube set at $2\theta = 15.6°$, no reflection occurs because the scattered rays from successive (100) planes are out of phase with each other, one plane of atoms exactly canceling the contribution from its neighbor. If the θ-angle of the crystal surface is increased to 15.8°, the alternate planes reinforce each other and a very strong reflection is detected by the counter tube at $2\theta = 31.6°$. This is the second-order reflection of (100), or the first-order reflection of the (200) planes, $d = 2.82$ A. The (300) planes also do not reflect, but the (400) planes reflect at $2\theta = 66.3°$, although less strongly than the (200) planes. If other sections were sawed parallel to (111) with $d = 3.26$ A, and to (220) with $d = 1.99$ A, reflections of different intensities would have been obtained with the counter tube set at 27.4° and 45.5°, respectively.

If longer wavelength x-rays had been used, the reflections would have occurred at correspondingly larger angles and the dispersion would have been greater, since the 2θ-angle at which the reflection from a given set of planes of spacing d occurs is proportional to $\lambda/\sin \theta$. For a given wavelength, larger d spacings appear at smaller angles Moreover, if the Bragg equation is differentiated, one obtains

$$\Delta\theta/\Delta d = -(\tan \theta)/d$$

which shows that the shift in line position $\Delta\theta$ due to a change of lattice spacing Δd increases as the tangent of the angle and reaches a maximum at $\theta = 90°$ (reflection angle $2\theta = 180°$), where $\tan \theta$ is infinite Hence the highest-angle lines in the back-reflection region of the pattern are the most sensitive to changes in the lattice spacings and consequently supply the most accurate data for the measurement of the unit cell dimensions

Many materials are not available in the form of large single crystals, and moreover it is impractical to obtain all the x-ray reflections from single crystals for identification purposes. If the sample does

the structure and there is little, if any, change in the dimensions. See ALLOY STRUCTURES.

Precision measurements of unit cell dimensions, or lattice parameters, are required for many types of solid state studies. For example, measurements at two known temperatures permit the calculation of the coefficient of thermal expansion. The data have also been used to compute the atomic weight of certain elements with a precision approaching that obtained by chemical methods. When the density of the crystals is known, the data may be used to compute the molecular weight.

In phase diagram analyses, the x-ray powder method is an extremely valuable tool to determine the limits of solubility and to identify the phases that occur. In solid state chemical reactions, chemical analysis is of little help because the bulk chemical composition of the starting constituents is the same as the desired end product. X-ray diagrams are therefore used to follow the course of the reaction and to determine if the desired end product has been formed.

Analysis of physical properties. When a substance is strained or plastically deformed, the x-ray lines broaden, and when the strain is removed by annealing, the lines return to their original sharpness. This is illustrated by the diffractometer recordings in Fig. 5. The upper recording shows the pattern of a sample of $BaTiO_3$ which has been strained by crushing, while the lower pattern shows the same substance in an unstrained state. X-ray patterns can thus be used to follow the course of heat treatment or other processes used to remove strains.

If the crystallites do not have a completely random orientation, the line shapes and relative intensities will change accordingly. For example, in rolling thin sheets of metal or in drawing wire, the crystallites align themselves in special ways, depending on the mechanical conditions of the process. The special or preferred orientation of the crystallites gives x-ray diffraction patterns which may be markedly different from those of the random crystallites. Similar conditions may arise in electroplating, where the plating conditions or the substrate may cause the crystallites to have a nonrandom orientation. Comparison of the random and oriented x-ray patterns shows the degree of orientation in the sample.

When the crystallites are very small, for instance, < 1 micron, the x-ray lines broaden by an amount which increases with decreasing size. Comparisons with the line breadths of samples of larger crystallite sizes may lead to a measure of the average crystallite size in the specimen. [W P]

Bibliography: L. V. Azaroff and M. J. Buerger, *The Powder Method in X-Ray Crystallography*, 1958; C. S. Barrett, *Structure of Metals*, 2d ed., 1952; E. U. Condon and H. Odishaw (eds.), *Handbook of Physics*, 1958; H. P. Klug and L. E. Alexander, *X Ray Diffraction Procedures*, 1954; W. Parrish, Advances in x ray diffractometry of clay minerals, *Proc. Seventh Natl. Conf. on Clays and Clay Minerals*, 1959; W. Parrish and T. R. Kohler, The use of counter tubes in x ray analysis, *Rev. Sci. Instr.*, 27(10):795–808, 1956.

X-ray tube

An electronic device used for the generation of x-rays. X-rays are produced in the x-ray tube by accelerating electrons to a high velocity by an electrostatic field and then suddenly stopping them by collision with a solid body, the so-called target, interposed in their path. The x-rays radiate in all directions from the spot on the target where the collisions take place. The x-rays are due to the mutual interaction of the fast moving electrons with the electrons and positively charged nuclei which constitute the atoms of the target. Depending upon the method used in generating the electrons, x-ray tubes may all be classified in two general groups, gas tubes and high-vacuum tubes. See X-RAY(S), PHYSICAL NATURE OF.

Gas x-ray tubes. In gas tubes electrons are freed from a cold cathode by positive ion bombardment. For the existence of the positive ions a certain gas pressure is required without which the tube will allow no current to pass. In the earliest gas tube the electrons liberated by positive ion bombardment from a flat aluminum cathode were emitted in a direction perpendicular to the cathode surface. The electrons traveled in straight lines until they impinged upon the glass end wall of the tube, where x-rays were generated.

Many designs of gas tubes have been built for useful application, particularly in the medical field. Metals such as platinum and tungsten, have been placed in the path of the electron beam to replace glass as the target. Concave metal cathodes are used to focus the electrons on a small area of the metal target and increase the sharpness of the resulting shadows on the fluorescent screen or the photographic film. Figure 1 shows one form of commercial gas x-ray tube.

The useful life of a gas x-ray tube is dependent on maintaining a constant gas pressure. In operation, the gas pressure in the tube will change, either rising or falling, depending on the particular operating conditions and on the past history of the

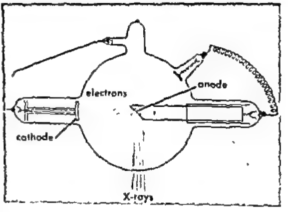


Fig. 1 A commercial model of gas x-ray tube

One of the most important characteristics of powder patterns is that isostructural substances (that is, substances with the same crystal structure) give similar diffraction patterns. The distribution of spacings and intensities of corresponding lines are nearly the same. The atoms in diamond, silicon, and sphalerite (ZnS) are all tetrahedrally bonded to their neighbors and all have the diamond-type structure. The differences in the spacings and relative intensities are dependent upon the atomic sizes and their x-ray-scattering power. Charts showing the powder patterns of structure types in the cubic, tetragonal, and hexagonal systems have been compiled, and it is often possible to identify the structure type by visual inspection of the patterns.

Many substances may occur in two or more crystal structures, that is, they may have polymorphic forms but the same chemical composition. For example, sphalerite and wurtzite are both ZnS. Such forms may be caused by slight differences in chemical preparation, different heat treatments, and other factors.

When two or more substances are present in the sample, the pattern of each substance appears independent of the others. This makes the identifications more difficult, but mixtures containing as many as six substances have been successfully analyzed. It is also possible to make quantitative analyses of the mixtures by comparing the relative intensities of one or more principal lines of each substance. Usually some reference standards of known chemical compositions are prepared to facilitate the interpretation.

Amorphous substances. Amorphous materials such as glasses and liquids give patterns which consist of only a few broad lines superimposed on a continuous background. Patterns of various amorphous substances closely resemble each other and hence the method is impractical for identification. On the other hand, since the two types of patterns are so different, the method is ideally suited to distinguish between crystalline and amorphous substances and to determine the degree of crystallinity of substances between the two extremes. The study of the progress of devitrification (crystallization) of a glass and similar problems are frequently accomplished with x-rays. Certain chemical and structural properties of liquids might be ascertained from the diffraction patterns when the liquids are frozen to form crystals.

Solid solutions. There are also many smaller changes in the x-ray pattern which may reveal important information. In substitutional solid solutions, for example, atoms of different elements may substitute for one another and occupy the same relative positions as in the pure metals (*see* SOLID SOLUTION). The substitutions of solute atoms occur on the same lattice sites occupied by the solute atoms, but are randomly distributed. If the atoms are of different size, the average unit cell size will change accordingly. In simple cases, it is possible to determine the chemical composition of intermediate members by measuring the unit cell dimensions because there is often a nearly linear relationship between the two. In interstitial solid solutions, atoms are added to the empty spaces in

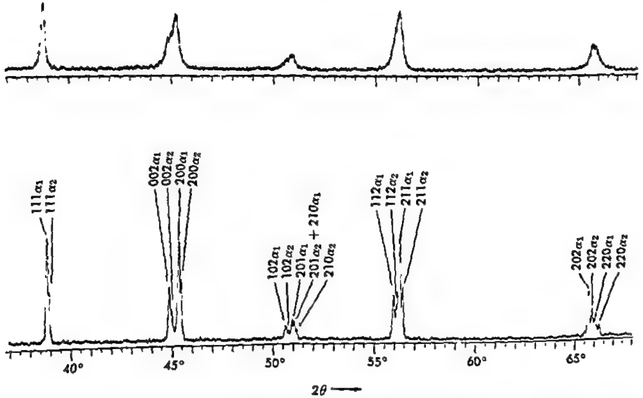


Fig. 5. Diffractometer patterns of strained $BaTiO_3$ (above) and unstrained $BaTiO_3$ (below). Strain has broadened the diffracted lines (*Modified from Trans Instr. and Meas. Conf., Stockholm, 1952*)

section x-ray tubes must preferably be operated at voltages below which field currents cannot be produced.

There are, however, some highly specialized applications of x-ray tubes for which the tube is designed specifically to produce and use field currents to generate x-rays. In order to build x-ray tubes that operate stably and continuously at very high voltages, the multisection principle is usually employed. These tubes are made with many intermediate sections between the cathode and anode sections. The voltage applied across each section of this multisection tube is always less than that at which field currents will be produced. The electrons emitted from the cathode are accelerated by the voltage applied across each intermediate section. The sum of all of these sectional voltages determines the voltage rating of the multisection x-ray tube. By this procedure, x-ray tubes can be built to generate x-rays at many million volts (Mev). Figure 5 is a line drawing of a commercial 2-Mev x-ray tube built according to this principle.

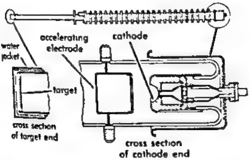


Fig. 5. A 2,000,000-volt multisection, hot-cathode, high-vacuum x-ray tube.

The toroidal electromagnetic type of hot-cathode high-vacuum x-ray tube, first successfully built by D. W. Kerst, permits the generation of x-rays at energy levels of many million electron volts without encountering the high-voltage insulation problem that is present in more conventional types of x-ray tubes. This x-ray tube consists of a circular "doughnut" type of high-vacuum envelope housing the hot-cathode and the x-ray target. This tube is designed to guide and accelerate electrons in a circular orbit in a machine called the betatron. In this toroidal shaped betatron tube (Fig. 6) the electrons are injected and energized to many millions of volts before they strike the target to produce x-rays. The average electron beam current is very small, but x-ray tubes of this type have been built to generate x-rays at energies ranging from 1 to over 100 Mev. The synchrotron is a circular accelerator related to the betatron and contains a similar doughnut-shaped x-ray tube. See PARTICLE ACCELERATOR.

The multisection linear accelerator type x-ray tube is another hot-cathode high-vacuum device that is capable of generating intense x-rays over a wide range of energy levels ranging from 1 to many Mev. It consists (Fig. 7) of an electron gun and a copper pipe wave guide in which the electrons are accelerated by the axial electric field produced by a high-frequency oscillator. At the output end of the wave guide, the electrons strike a target and generate x-rays. X-ray tubes of this type have been built for operation at 5–15 Mev and higher, and with average currents of the order of 25 microamperes or more.

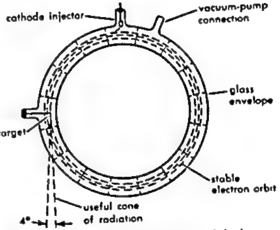


Fig. 6. Betatron "doughnut" type of high-vacuum x-ray tube. (General Electric Company)

Applications. Many special forms of x-ray tubes, following the basic patterns already outlined, have been built for application in medicine, industry, and fundamental science. They vary in size from a few inches to many feet in length. X-ray tubes are used as an aid in medical diagnosis and in therapeutic treatment. They are used in the nondestructive testing of materials throughout all industry. In both fundamental and applied science they find wide use in crystal-diffraction work, chemical analysis by x-ray spectra and absorption, and research on atomic structure.

To provide maximum use in these wide fields of application, particular design features are built into the tubes to meet the special requirements. Some of these special requirements are low or high currents, low- or high-voltage x-rays, large or small focal spot, high x-ray beam intensity for short or long intervals, and choice of x-ray target material to generate a particular quality of x-ray spectra. X-ray and electrical protection must be ensured.

The over-all capacity of an x-ray tube is determined principally by the target material, the area

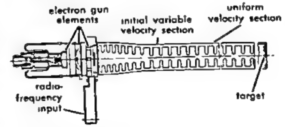


Fig. 7. A linear accelerator type of hot-cathode, high-vacuum x-ray tube.

tube. The pressure will usually be reduced by operating the tube intermittently with small currents. When the pressure becomes too low for satisfactory operation, it can be raised again by various methods, prolonging the useful life of the tube. The pressure can be raised by heating the glass bulb; by heating a suitable chemical placed in a side tube; or by diffusing gas through a thin-walled tube of platinum, palladium, or unglazed porcelain joined to the tube envelope. The pressure may often be raised when operating the tube continuously, as in therapeutic work. The gas pressure is often regulated automatically. Many forms of automatic osmoregulators were designed and used to admit or remove gas to maintain the desired operating pressure in the gas x-ray tube.

The size of the bulb in commercial tubes was increased as the power input was increased, to reduce the pressure change in operation, and to reduce the local heating of the glass envelope resulting from bombardment of electrons reflected from the focal spot. The aluminum cathodes were made heavier to withstand the increased positive ion bombardment. The thin metal targets were also replaced by a heavier mass of metal. Targets now consist of two main parts: a refractory metal face, such as platinum, to take the direct impact of the electron beam, and a heavy backplate of a good heat-conducting material, such as copper, to conduct the heat away from the focal spot and store it temporarily.

Early in the development of x-ray tubes metals of high atomic weight were known to be the most efficient x-ray generators. W. K. Röentgen had used both aluminum and platinum as targets. W. D. Coolidge and others later used thorium and uranium with increased efficiency The principal properties desired for a suitable target material are (1) high atomic number to give best x-ray efficiency, (2) high melting point and high thermal conductivity to permit maximum energies for a given size of focal spot, and (3) low vapor pressure to reduce the rate of evaporation of the metal on the walls of the glass envelope. Ductile tungsten as developed by Coolidge was found to combine these desired properties of an x-ray target to the greatest degree

High-vacuum x-ray tubes. The operational difficulties and erratic behavior of gas x-ray tubes are inherently associated with the gas itself and the positive ion bombardment that takes place during operation The high-vacuum x-ray tube eliminates these difficulties by using other means of emitting electrons from the cathode.

Coolidge made a high-vacuum x-ray tube with a hot tungsten-filament cathode and a solid tungsten target. This hot-cathode high-vacuum type of x-ray tube permitted stable and reproducible operation with relatively high voltages and large masses of metals. The vacuum was so good that positive ions did not play either an essential or a harmful role in the tube operation. The earliest commercial form of the hot-cathode high-vacuum type of x-ray tube (Fig. 2) utilizes a solid tungsten target. The independent relation of the current to the impressed

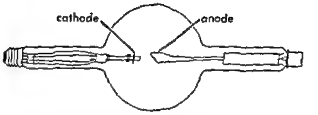


Fig. 2. Early commercial model of single-section, hot-cathode, high-vacuum x-ray tube.

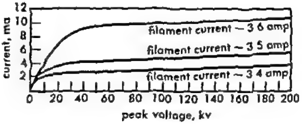


Fig. 3. Curves showing relation of current to voltage in hot-cathode high-vacuum x-ray tube.

voltage in a tube of this type is shown in Fig. 3 The different curves are for different filament temperatures and show that, over the operating range of x-ray voltages, the discharge current is practically independent of voltage. One typical form of a modern commercial hot-cathode high-vacuum x-ray tube (Fig. 4) is built with a liquid-cooled, copper-backed tungsten target, which operates over a wide range of energy ratings and is capable of rectifying its own current.

The inherent advantages of the hot-cathode high-vacuum tube over the gas tube are (1) flexibility—the voltage and current may be varied independently, (2) stability—permitting more accurate reproducibility of results; (3) small size; (4) operation—it can be operated directly from a transformer, making possible a very simple unit; and (5) long life

High-voltage x-ray tubes. The operating voltage of a single-section hot-cathode x-ray tube is limited by the field current pulled from the cathode by the electrostatic field (see FIELD EMISSION). If produced in an x-ray tube, field currents may cause erratic and uncontrollable operating behavior Single-

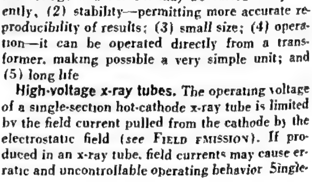


Fig 4 A modern single-section, hot-cathode, high-vacuum x-ray tube with a liquid-cooled copper-back tungsten target.

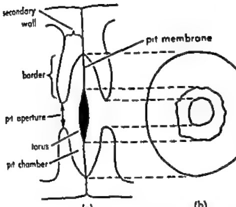


Fig 3 Diagrams of bordered pit-pairs of *Pinus*. (a) Sectional view (b) Face view. Torus formed by thickening of primary walls (From K. Esau, *Plant Anatomy*, Wiley, 1953)

Tracheids Tracheids are elongate, spindle-shaped cells, lacking protoplasm at maturity, and having secondary walls laid in various thicknesses and patterns over the primary wall (Fig 2). The secondary wall may consist of stretchable annular or helical (spiral) bands, or of a nonextensible continuous expanse interrupted only at the bordered pits

Vessels. Vessel members differ from tracheids in having perforations, usually in the end walls common to superposed cells A longitudinal series of vessel members constitutes a vessel. Although each varies markedly, tracheids are usually twice to three times as long as vessel members, for example, 1–1.5 mm versus 0.5 mm in secondary wood. Vessels, however, may be from many centimeters to even meters in length

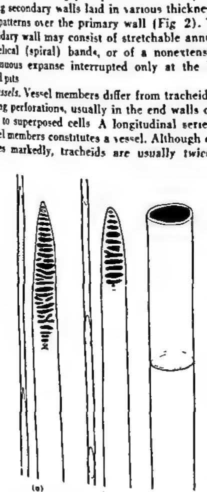


Fig 4 Complete vessel members and parts of single vessel members with perforation plates on end walls (perforations in black). (a) Long scalariform perforation plate (b) Shorter scalariform (c) Simple

In wood, pits are interruptions or recesses in the secondary wall, and are usually paired in common walls of contiguous cells. In bordered pits (Fig. 3) the secondary wall overarches the pit membrane (intercellular substance and two primary walls of contiguous cells); the overarch is absent in simple pits. The borders may be well developed or barely perceptible. The pit apertures (openings into the cell lumen) in bordered pits may be round to slitlike openings. Vertical series of transversely elongated pits are described as scalariform pitting Round or oval pits may appear singly, or in horizontal (opposite pitting) or oblique (alternate pitting) series.

Perforations result from disintegration of the primary wall in the ends of vessel members. A perforation plate (Fig. 4) is the wall area bearing one or more perforations. The most primitive (scalariform) perforation plates are strongly oblique in position and have numerous transversely placed perforations (over 200). The most specialized have a single large (simple) perforation in a transverse wall.

Fibers Fibers differ from tracheids in being more slender, having simple pits and thicker walls in proportion to diameter, and in often retaining protoplasm Fibers are thought to have evolved from tracheids, and intergrading forms between them are called fiber-tracheids Divisions of the protoplast of living fibers result in septate fibers.

Parenchyma The term *parenchyma* usually refers to the vertically arranged living cells of the xylem, although rays are also composed chiefly of parenchyma Depending upon their origin, parenchyma cells may occur singly or as members of a strand The cells may or may not have secondary walls, typically have simple pits, usually store starch, and may contain crystals, tannins, or other substances

Primary and secondary xylem. *Xylem tissues* arise in later stages of embryo development of a given plant and are added to by differentiation of cells derived from the apical meristems of roots and stems Growth and differentiation of tissues derived from the apical meristem provide the primary body of the plant, and the xylem tissues formed in it are called primary Secondary xylem, when present, is produced by the vascular cambium (*see* MERISTEM, LATERAL).

Primary xylem Protoxylem is that part of the primary xylem formed while the organ is still elongating and generally includes cells with extensible secondary wall thickening These cells are frequently torn in vigorously growing plants. Metaxylem is formed after growth in length is nearly completed and has cells mostly with nonextensible secondary walls Parenchyma cells, with thin or thick walls, occur in both protoxylem and metaxylem, and, depending upon the species, vessels, tracheids, and fibers all may be present.

of the focal spot, the duration of the energy applied and the temperature of the target during the time the electron energy is applied. The modern x-ray tube is a precision tool of great stability and flexibility, capable of controlled operation with currents and voltages of any desired magnitude, and generating x-rays over a wide range of energy and intensity. *See* NUCLEAR RADIATION, BIOLOGY; RADIATION CHEMISTRY; RADIOGRAPHY; RADIOGRAPHY OF METALS; X-RAY FLUORESCENCE ANALYSIS.

[E.E.C.]

Bibliography: E. E. Charlton, High voltage sources for industrial radiography—past and present, and future trends, *J. Soc. Nondestructive Testing,* 13(1):13–22, 1955; O. Glasser (ed), *The Science of Radiology,* 1933

X-unit

A unit of length, approximately 10^{-11} cm, used for measuring x-ray and γ-ray wavelengths and crystal dimensions. It is also called the Siegbahn unit, in honor of M. Siegbahn who, in 1919, noted that relative x-ray wavelengths and crystal dimensions could be measured more precisely than their absolute values. Hence he proposed for the measurement of these quantities a new unit of length, the X-unit (XU), defined such that the grating constant of a calcite crystal for first order diffraction was precisely 3029.04 XU.

X-ray diffraction from crystals is equivalent to diffraction from a three-dimensional grating (*see* X-RAY DIFFRACTION). From such measurements one can determine grating constants (dimensions of a unit cell) in terms of wavelengths or vice versa. The grating constants can also be calculated from Avogadro's number, molecular weight, and crystal density. On the basis of the best data available before 1930, such calculations gave a value of approximately 3029×10^{-11} cm for the grating constant of calcite

Subsequently, absolute measurements of x-ray wavelengths with ruled gratings have been made by J. A. Bearden, E. Bäcklin, M. Söderman, and F. Tyrén, these indicate that the XU exceeds 10^{-11} cm by about two parts per thousand. In 1947 W L. Bragg, after consultation with authorities in the field, recommended the conversion factor, $1\ XU = (1.00202 \pm .00003) \times 10^{-11}$ cm. Recent evaluations of atomic constants lead to values of this factor in substantial agreement with the directly measured one quoted here. *See* ATOMIC CONSTANTS.

[J A.B., J.S.T.]

Xylem

The principal water-conducting tissue, which also serves as the chief supporting system. This tissue and the closely associated phloem constitute the vascular system of vascular plants Xylem is composed of various kinds of cells, living or nonliving. The structure of these cells differs with their functions, but characteristically all have a rigid and enduring cell wall that is well preserved in fossils.

Kinds of xylem cells. In terms of their functions, the kinds of cells in xylem are those related principally to conduction and support, tracheids; to conduction, vessel members; to support, fibers; and to food storage, parenchyma (Fig. 1). Vessel members and tracheids are often called tracheary elements. The cells in each of the four categories vary widely in structure.

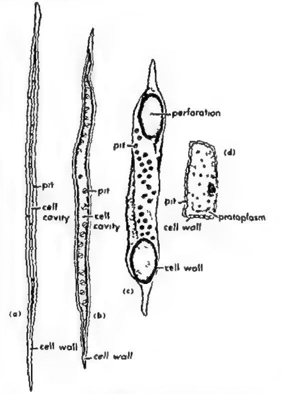


Fig 1. Xylem cell types (a) Wood (xylem) fiber. (b) Tracheid (c) Vessel member (d) Xylem parenchyma cell (From H J Fuller and O. Tippo, College Botany, rev. ed., Holt, 1954)

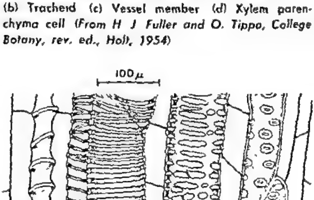


Fig. 2. Parts of tracheary elements and associated parenchyma from stem of Aristolochia, as seen in longitudinal section. The secondary walls are stippled Earliest part of xylem at left. (From K Esau, Plant Anatomy, Wiley, 1953)

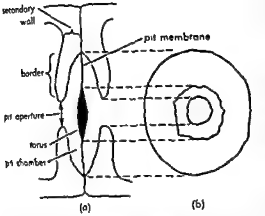


Fig 3 Diagrams of bordered pit-pairs of *Pinus* (a) Sectional view (b) Face view. Torus formed by thickening of primary walls (From K. Esau, *Plant Anatomy*, Wiley, 1953)

Tracheids Tracheids are elongate, spindle-shaped cells, lacking protoplasm at maturity, and having secondary walls laid in various thicknesses and patterns over the primary wall (Fig 2) The secondary wall may consist of stretchable annular or helical (spiral) bands, or of a nonextensible continuous expanse interrupted only at the bordered pits

Vessels Vessel members differ from tracheids in having perforations, usually in the end walls common to superposed cells A longitudinal series of vessel members constitutes a vessel Although each varies markedly, tracheids are usually twice to three times as long as vessel members, for example, 1–1.5 mm versus 0 5 mm in secondary wood. Vessels, however, may be from many centimeters to even meters in length.

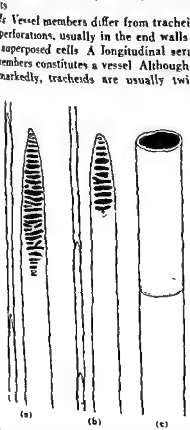


Fig 4 Complete vessel members and *parts* of single vessel members with perforation plates on end walls (perforations in black) (a) Long scalariform perforation plate, (b) Shorter scalariform, (c) Simple

In wood, pits are interruptions or recesses in the secondary wall, and are usually paired in common walls of contiguous cells. In bordered pits (Fig. 3) the secondary wall overarches the pit membrane (intercellular substance and two primary walls of contiguous cells), the overarch is absent in simple pits The borders may be well developed or barely perceptible. The pit apertures (openings into the cell lumen) in bordered pits may be round to slitlike openings. Vertical series of transversely elongated pits are described as scalariform pitting. Round or oval pits may appear singly, or in horizontal (opposite pitting) or oblique (alternate pitting) series.

Perforations result from disintegration of the primary wall in the ends of vessel members A perforation plate (Fig. 4) is the wall area bearing one or more perforations. The most primitive (scalariform) perforation plates are strongly oblique in position and have numerous transversely placed perforations (over 200). The most specialized have a single large (simple) perforation in a transverse wall

Fibers Fibers differ from tracheids in being more slender, having simple pits and thicker walls in proportion to diameter, and in often retaining protoplasm Fibers are thought to have evolved from tracheids, and intergrading forms between them are called fiber-tracheids. Divisions of the protoplast of living fibers result in septate fibers.

Parenchyma The term parenchyma usually refers to the vertically arranged living cells of the xylem, although rays are also composed chiefly of parenchyma Depending upon their origin, parenchyma cells may occur singly or as members of a strand The cells may or may not have secondary walls, typically have simple pits, usually store starch, and may contain crystals, tannins, or other substances.

Primary and secondary xylem. Xylem tissues arise in later stages of embryo development of a given plant and are added to by differentiation of cells derived from the apical meristems of roots and stems Growth and differentiation of tissues derived from the apical meristem provide the primary body of the plant, and the xylem tissues formed in it are called primary Secondary xylem, when present, is produced by the vascular cambium (*see* MERISTEM, LATERAL).

Primary xylem Protoxylem is that part of the primary xylem formed while the organ is still elongating and generally includes cells with extensible secondary wall thickening. These cells are frequently torn in vigorously growing plants. Metaxylem is formed after growth in length is nearly completed and has cells mostly with nonextensible secondary walls Parenchyma cells, with thin or thick walls, occur in both protoxylem and metaxylem, and, depending upon the species, vessels, tracheids, and fibers all may be present.

Secondary xylem. Secondary xylem is composed of two interpenetrating systems, the horizontal (ray) and vertical (axial). Wide variation occurs in secondary wood, mostly because of inherent differences among species, partly because of environmental influences. The differences in rate of evolutionary specialization of the different kinds of cells as seen in any one species make possible the positive identification of nearly *all* kinds of wood. Variations include differences in types of rays and their distribution; kinds, numbers, arrangement, and sizes of vessels, tracheids, and fibers; types of growth layers; character of sapwood and heartwood; kinds of knots; distribution of parenchyma; and other features, such as odor. Many of these variations bear on the strength and usefulness of wood.

Rays may be one to many cells in width as seen in transections of the organ in which they occur, may be composed of cells alike or unlike in form, are variable in length radially, and differ in height, for example, high in oak, low in maple. Evolutionary specialization has involved height and width and even loss of rays, and changes in dimensions of the ray cells themselves. Activities of ray cells include the formation of tyloses—protrusions of parenchyma cells into the lumen of tracheary elements—and the secretion of variable quantities of gummy substances into neighboring cells.

A growth layer is the increment of secondary xylem produced in one growing season (Fig. 5). It is called an annual layer—or annual ring when seen in transection—if a single such season occurs each year. The boundaries between successive growth rings are identifiable largely because the xylem cells in the later part of one season (late or summer wood) are much smaller than the first ones produced in the next season's growth (early or spring wood). In ring-porous woods, the vessels are much larger in the early wood than in the late wood of a given growth layer. The vessels are of approximately equal size throughout in diffuse-porous woods (Figs. 6, 7).

Most knots are bases of twigs or branches buried, either alive (tight knot) or dead (loose knot), by cambial activity of the mother branch (Fig 8). Their size depends largely upon the age at which they were buried. In lumber, round knots represent transections of the encased branch and spike knots, longitudinal sections

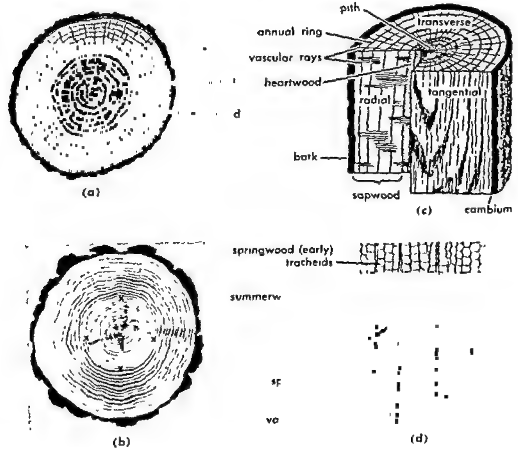


Fig 5 (a) Cross section of oak branch showing bark, heartwood (dark inner portion), sapwood (light portion between bark and heartwood), vascular rays, and annual rings. (b) Cross section of loblolly pine trunk showing effect of light upon annual ring thickness. The annual rings inside the xs were formed while the young tree was densely shaded by older trees. The annual rings outside the xs were formed after the older trees had been removed and the young pine tree received abundant sunlight (c) Three-dimensional appearance of a log (d) Highly magnified portion of transverse section of pine wood showing portions of two annual rings (From H J Fuller and O Tippo, *College Botany*, rev ed., Holt, 1954)

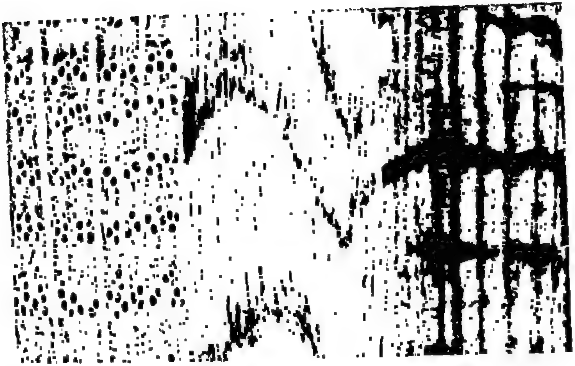


Fig 6 Red oak wood (a ring-porous hardwood) (a) Transverse section (b) Tangential section. (c) Radial section (From H J. Fuller and O. Tippo, College Botany, rev ed, Holt, 1954)

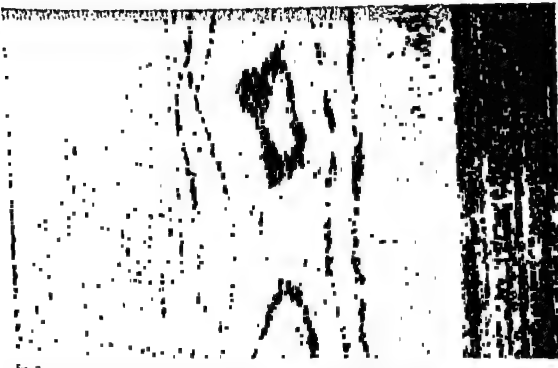


Fig 7. Yellow birch wood (a diffuse-porous hardwood) (a) Transverse section (b) Tangential section (c) Radial section (light portion is sapwood, dark portion is heartwood) (From H. J. Fuller and O Tippo, College Botany, rev. ed, Holt, 1954)

Secondary xylem. Secondary xylem is composed of two interpenetrating systems, the horizontal (ray) and vertical (axial). Wide variation occurs in secondary wood, mostly because of inherent differences among species, partly because of environmental influences. The differences in rate of evolutionary specialization of the different kinds of cells as seen in any one species make possible the positive identification of nearly all kinds of wood. Variations include differences in types of rays and their distribution; kinds, numbers, arrangement, and sizes of vessels, tracheids, and fibers; types of growth layers; character of sapwood and heartwood; kinds of knots; distribution of parenchyma; and other features, such as odor. Many of these variations bear on the strength and usefulness of wood.

Rays may be one to many cells in width as seen in transections of the organ in which they occur, may be composed of cells alike or unlike in form, are variable in length radially, and differ in height, for example, high in oak, low in maple. Evolutionary specialization has involved height and width and even loss of rays, and changes in dimensions of the ray cells themselves. Activities of ray cells include the formation of tyloses—protrusions of parenchyma cells into the lumen of tracheary elements—and the secretion of variable quantities of gummy substances into neighboring cells.

A growth layer is the increment of secondary xylem produced in one growing season (Fig. 5). It is called an annual layer—or annual ring when seen in transection—if a single such season occurs each year. The boundaries between successive growth rings are identifiable largely because the xylem cells in the later part of one season (late or summer wood) are much smaller than the first ones produced in the next season's growth (early or spring wood). In ring-porous woods, the vessels are much larger in the early wood than in the late wood of a given growth layer. The vessels are of approximately equal size throughout in diffuse-porous woods (Figs. 6, 7).

Most knots are bases of twigs or branches buried, either alive (tight knot) or dead (loose knot), by cambial activity of the mother branch (Fig. 8). Their size depends largely upon the age at which they were buried. In lumber, round knots represent transections of the encased branch and spike knots, longitudinal sections.

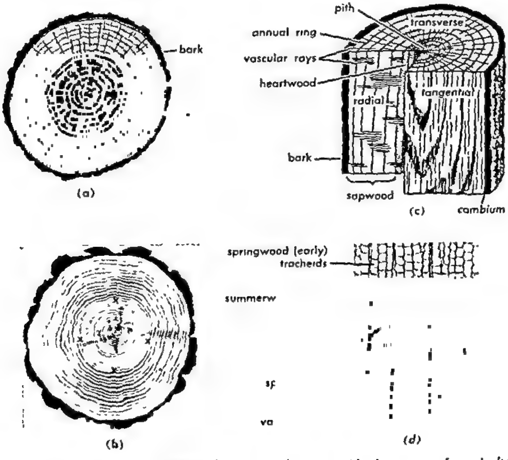


Fig. 5. (a) Cross section of oak branch showing bark, heartwood (dark inner portion), sapwood (light portion between bark and heartwood), vascular rays, and annual rings. (b) Cross section of loblolly pine trunk showing effect of light upon annual ring thickness. The annual rings inside the xs were formed while the young tree was densely shaded by older trees. The annual rings outside the xs were formed after the older trees had been removed and the young pine tree received abundant sunlight. (c) Three-dimensional appearance of a log. (d) Highly magnified portion of transverse section of pine wood showing portions of two annual rings. (From H. J. Fuller and O. Tippo, *College Botany, rev. ed.*, Holt, 1954)

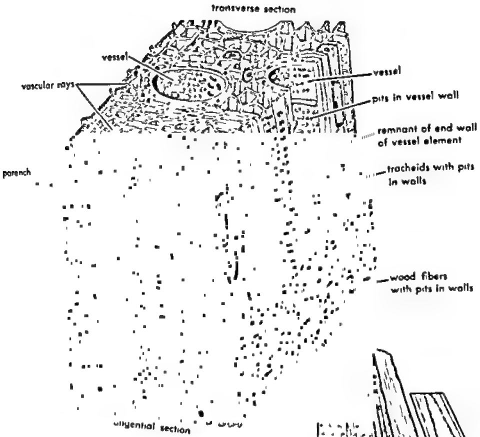


Fig 10 Microscopic structure of oak (angiosperm) wood (three-dimensional) (From H J Fuller and O Tippo, College Botany, rev ed., Holt, 1954)

Softwood and hardwood. In the trade, softwood is a name for xylem of gymnosperms (conifers) and hardwood for xylem of angiosperms The terms do not refer to actual hardness of the wood Woods of gymnosperms are generally composed only of tracheids, wood parenchyma, and small rays, but differ in detail (Fig 9) Resin ducts are present in many softwoods. Woods of angiosperms show extreme variation in both vertical and horizontal systems, but with few exceptions have vessels (Fig 10)

The grain, texture, and figure of wood reflect gross differences in the arrangement, orientation, and kinds of cells in xylem and often are useful in identification (Fig 11). These differences affect both beauty and strength of wood Specific gravity also influences strength, as do the size arrangement, and number of vessels, fibers parenchyma cells, and rays Heartwood—the older wood—is no stronger than the younger sapwood, although often it is more durable Transformation of sapwood into heartwood involves partial loss of water, death of living cells, production of various substances—often colored, and frequently the appearance of tyloses *See* CYTOLOGY; FOREST AND FORESTRY, PLANT TISSUE SYSTEMS [V I C]

Bibliography: *See* PLANT ANATOMY

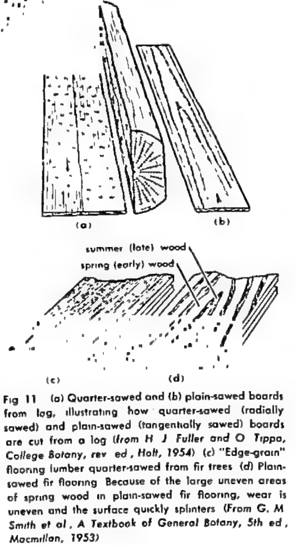


Fig 11 (a) Quarter-sawed and (b) plain-sawed boards from log, illustrating how quarter-sawed (radially sawed) and plain-sawed (tangentially sawed) boards are cut from a log (from H J Fuller and O Tippo, College Botany, rev ed, Holt, 1954) (c) "Edge-grain" flooring lumber quarter-sawed from fir trees (d) Plain-sawed fir flooring Because of the large uneven areas of spring wood in plain-sawed fir flooring, wear is uneven and the surface quickly splinters (From G. M Smith et al, A Textbook of General Botany, 5th ed, Macmillan, 1953)

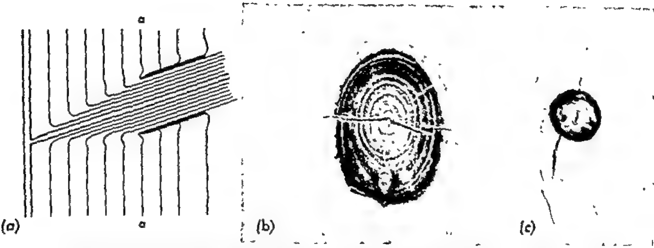


Fig. 8. (a) Diagram illustrating the burial of the base of a branch by secondary growth. The portion of the branch to the left of the line *a–a* was alive when buried, and its growth rings are continuous with those of the trunk. The portion to the right of *a–a* was dead when buried, and the growth rings of branch and trunk are not continuous. Pith finely stippled, bark coarsely stippled. (b) Hard pine board showing tight knot. (c) Hard pine board showing loose knot. *(From C. L. Wilson and W. E. Loomis, Botany, rev. ed., Dryden, 1957)*

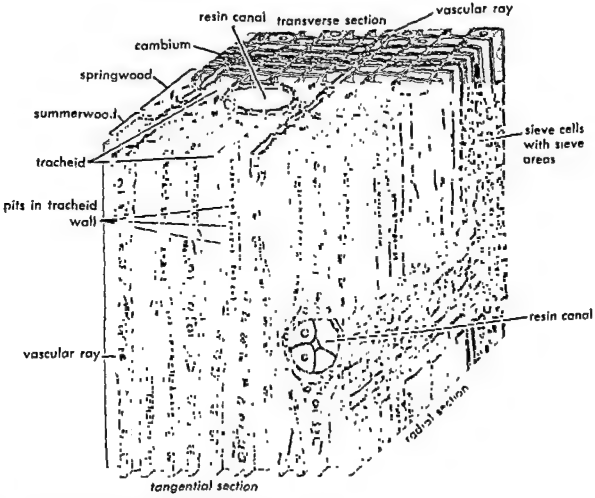


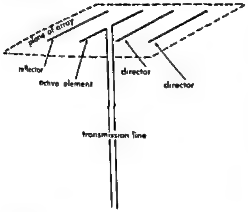
Fig. 9. *Microscopic structure of a portion of a pine (gymnosperm) stem (three-dimensional) (From H. J. Fuller and O. Tippo, College Botany, rev. ed., Holt, 1954)*

Y *Yagi antenna to Yttrium*

Yagi antenna

A common form of *directional* radio antenna. The Yagi antenna consists of an array of parallel coplanar linear conductors which employs parasitic conductors called directors on one side of the exciter, or driven conductor, so as to improve transmission in that direction. Directivity is enhanced by a reflector on the opposite side (see illustration).

Correct operation depends on the relative phases of currents in the conductors, which in turn are controlled by conductor length, the reflector being

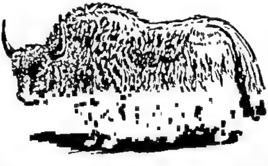


Yagi antenna

approximately at resonant (half-wave) length, the directors being 5 or 10% shorter. Typical spacing between conductors is ⅕ wavelength. However, there is considerable latitude in lengths and spacings, subject to experimental adjustment, since length and spacing are interdependent. Many directors have been used, but an increase in number narrows bandwidth. The antenna is very useful where moderate gains are needed, where economy is important, and where narrow band operation is no handicap. Increased gain can be achieved by using Yagi antennas in array. *See* ANTENNA (AERIAL)
[J.C.S.]

Yak

A bovine, *Bos grunniens*, of the high, cold portions of central Asia, particularly associated with Tibet. Yaks have short curved horns, a hump similar to that of the American bison between the shoulders, and are covered with long hair. Yaks are both domestic and wild and can survive in rigorous areas where other domesticated animals perish. They supply hair for cloth, hides for leather, meat, and a rich milk from which butter and cheese are made. They are sturdy and sure-footed beasts of burden. *See* ARTIODACTYLA.
[J.D.B.]



The yak, *Bos grunniens;* larger than average domestic cattle and more solidly built. (From P. Martin Duncan, ed., *Cassell's Natural History*, Cassell)

Yaw indicator

An aircraft instrument used to measure the angle between the vertical plane containing the principal axis of the airplane and the air stream. Yaw indicators find common use only during flight testing.

A pivoted vane, mounted with its axis vertical, is installed exposed to the free air. The vane aligns itself with the direction of the air stream or relative wind. The position of the vane relative to a fixed reference line determines the yaw angle. This angular deflection positions a transducer, which transmits an electrical signal to an indicator.

The angle-of-attack indicator measures the inclination of a fore-and-aft reference line to the air stream. Angle-of-attack indicators find some application in stall warning. The instrument is similar to the yaw indicator except that its axis is mounted horizontally so that the vane may deflect vertically. *See* AIRCRAFT INSTRUMENTATION
[W.G.B.]

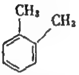
Yaws

An infectious disease of man caused by a spirochete *Treponema pertenue*. It is also known as frambesia and is largely confined to the tropics. Usually yaws is contracted in childhood by direct contact or from small flies feeding in succession on infected lesions and open wounds. No race or age possesses natural immunity.

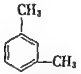
The causative spirochete, *Treponema pertenue*, is morphologically similar to *T. pallidum*. The general course of yaws resembles that of syphilis, but the individual yaws lesions tend to be larger and more persistent. The primary lesion or mother yaw

Xylene

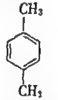
One of a group of colorless, liquid, aromatic hydrocarbons which are also called dimethylbenzenes. Although xylenes are still referred to as "coal-tar" hydrocarbons, nearly the entire supply is produced from petroleum by catalytic methods.



o-Xylene
bp 144.2°C
fp −25.3°C



m-Xylene
bp 139.1°C
fp −47.9°C



p-Xylene
bp 138.4°C
fp −13.3°C

Catalytic oxidation of *o*-xylene yields phthalic anhydride, previously available only by oxidation of naphthalene. Liquid-phase catalytic oxidation of *p*-xylene leads ultimately to terephthalic acid, an essential ingredient in an important polyester fiber and film. Isophthalic acid, formed by oxidation of *m*-xylene, is produced for use as a resin intermediate. For additional information, *see* OXIDATION PROCESS; PHTHALIC ACID.

A mixture of xylenes usually called "xylene" is used as a solvent for many organic substances. It is used in making lacquers and rubber cement. It is reported that prolonged breathing of air containing more than 200 parts per million of xylene vapor may be injurious to the health. *See* AROMATIC HYDROCARBON; BENZENE. [C.K.B.]

arm of the equivalent Y is equal to the product of the adjacents divided by the sum For example,

$$Z_1 = \frac{Z_A Z_B}{Z_A + Z_B + Z_C} = \frac{\text{product of adjacents}}{\text{sum}}$$

Y-Δ transformation. Let Z_1 be designated as the opposite of Z_C; Z_2 the opposite of Z_A; Z_3 the opposite of Z_B Also let $Z_1Z_2 + Z_2Z_3 + Z_3Z_1$ be called the sum of products. With this notation, Eqs. (10) through (12) state that each arm of the equivalent Δ section is equal to the sum of products divided by the opposite. For example,

$$Z_A = \frac{Z_1Z_2 + Z_2Z_3 + Z_3Z_1}{Z_1} = \frac{\text{sum of products}}{\text{opposite}}$$

[K Y.T.]

Year

Any of several units of time based on the revolution of Earth around the Sun. The tropical year, to which the calendar is adjusted, is the period required for the mean longitude of the Sun to increase 360°. Its duration is approximately 365.24220 mean solar days. It is also the period after which the seasons repeat themselves, and the unit in terms of which the second is officially defined.

The sidereal year, 365.25636 mean solar days in duration, is the average period of revolution of Earth with respect to a fixed direction in space

The anomalistic year, 365 25964 mean solar days in duration, is the average interval between successive closest approaches of Earth to the Sun. *See* Time.

[G.M.C]

Yeast

A collective name for those fungi which possess, der normal conditions of growth, a vegetative dy (thallus) consisting, at least in part, of sime. individual, single cells. In addition, the cells aking up the thallus may occur in pairs, in groups three, or in straight or branched chains consistg of as many as twelve or more cells Yeast plays large part in industrial fermentation processes ich as the production of ethanol, malt beverage, d wine and also in diseases of man, animals, and lants (*see* Distilled spirits; Ethyl alcohol; Malt beverage, Wine; Yeast, industrial)

Morphology. The shape and size of the individal cells of some species vary slightly, but in other pecies the cell morphology is extremely heteroeneous The shape of yeast cells may be spherical, lobose, ellipsoidal, elongate to cylindrical with ounded ends, more or less rectangular, pearhaped, apiculate or lemon-shaped, ogive or ointed at the ends, or tetrahedral The dimension of a spherical cell may vary from 2–10 microns (μ) n diameter The length of cylindrical cells is often 20–30 μ and in some cases, even greater.

The asexual multiplication of yeast cells occurs by a budding process, by the formation of crosswalls or fission, and sometimes by a combination of these two processes. Yeast buds are sometimes called blastospores When yeast reproduces by a fission mechanism, the resulting cells are termed arthrospores

In some genera of yeast true septate mycelium is formed, for example, in *Candida* and in *Trichosporon*. In the genus *Candida*, the mycelium, if present, remains intact, that is, it does not break up at the septa. In the genus *Trichosporon*, the mycelium breaks apart into arthrospores which at first are rectangular but often develop rounded ends when mature. Many yeasts produce a type of thallus which is termed a pseudomycelium. In a pseudomycelium the component cells are formed by elongation of buds and not by the formation of septa. Usually a group of characteristically arranged blastospores develops at the junctions of cells making up a pseudomycelium.

Yeasts are also characterized by certain macromorphological features. These are studied as slant

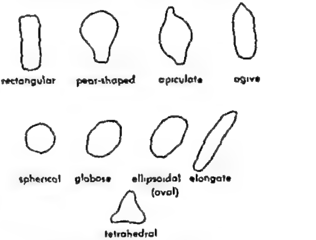


Fig 1 Various cell shapes of yeasts

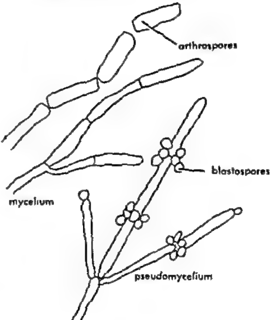


Fig 2 Drawing of a mycelium and a pseudomycelium

occurs commonly on the lower extremities. Generalized lesions are often large and fungating. Destructive bone lesions are common, as are painful incapacitating lesions of the soles of the feet (crab yaws). Central nervous system and cardiovascular involvement occur rarely if at all. Serological tests for syphilis are also positive in yaws, and the disease responds equally well to penicillin. Mass survey and treatment campaigns directed primarily to the reduction of infectious cases have been conducted with success in many countries. *See* PENICILLIN; SEROLOGY; SYPHILIS. [T.B.T.]

Y-delta transformations

Any Y-connected circuit may be transformed into an equivalent Δ-connected circuit or vice versa. This article presents methods of performing these transformations.

Figure 1*a* shows a Y section (also called a T section). Figure 1*b* shows a Δ section (or a π section). The Y section and the Δ section are sometimes classified as three-terminal networks

In the analysis and solution of a network, it may simplify the calculations to make the transformation of a Y section into a Δ section or vice versa By Y-Δ transformations, a passive linear bilateral network having two input terminals and two output terminals can be reduced to a simple Y section or a simple Δ section with input terminals *a-c* and output terminals *b-c* in Fig 1, as far as the currents at the input and output terminals are concerned.

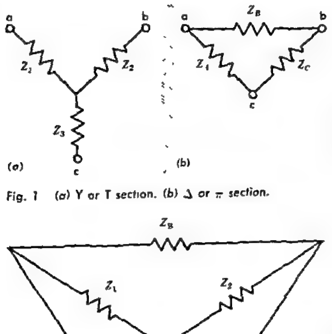


Fig. 1 (a) Y or T section. (b) Δ or π section.

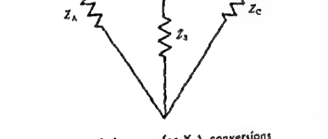


Fig. 2. General diagram for Y-Δ conversions

Such reductions are useful in analysis, since a linear bilateral network may be represented by a simple Y or T section.

A method of transforming a Δ section into an equivalent Y section, or vice versa, is to get the impedances across corresponding terminals *a-b*, *b-c* and *c-a* of the Y section equal to those of the Δ section; that is

$$(Z_{ab})_Y = (Z_{ab})_\Delta \tag{1}$$
$$(Z_{bc})_Y = (Z_{bc})_\Delta \tag{2}$$
$$(Z_{ca})_Y = (Z_{ca})_\Delta \tag{3}$$

Thus, for the notation and arrangements in Fig. 1,

$$Z_1 + Z_2 = \frac{Z_B(Z_A + Z_C)}{Z_A + Z_B + Z_C} \tag{4}$$

$$Z_2 + Z_3 = \frac{Z_C(Z_B + Z_A)}{Z_A + Z_B + Z_C} \tag{5}$$

$$Z_3 + Z_1 = \frac{Z_A(Z_C + Z_B)}{Z_A + Z_B + Z_C} \tag{6}$$

Δ-Y transformation. If Z_A, Z_B, and Z_C of the Δ section are given, then Z_1, Z_2, and Z_3 of the equivalent Y section can be found by solving equations (4) through (6) simultaneously and obtaining

$$Z_1 = \frac{Z_A Z_B}{Z_A + Z_B + Z_C} \tag{7}$$

$$Z_2 = \frac{Z_B Z_C}{Z_A + Z_B + Z_C} \tag{8}$$

$$Z_3 = \frac{Z_C Z_A}{Z_A + Z_B + Z_C} \tag{9}$$

Y-Δ transformation. If Z_1, Z_2, and Z_3 of the Y section are given, then Z_A, Z_B, and Z_C of the equivalent Δ section can be determined by solving equations (4) through (6) simultaneously and obtaining

$$Z_A = \frac{Z_1Z_2 + Z_2Z_3 + Z_3Z_1}{Z_2} \tag{10}$$

$$Z_B = \frac{Z_1Z_2 + Z_2Z_3 + Z_3Z_1}{Z_3} \tag{11}$$

$$Z_C = \frac{Z_1Z_2 + Z_2Z_3 + Z_3Z_1}{Z_1} \tag{12}$$

General rules. When Eqs (7) through (12) are applied, it must be understood that they hold for only the notation and arrangement of the impedances given in Fig 1 The following rules, with the aid of Fig. 2, might be helpful in determining the proper relations in the circuit for the new *Z*s when transforming from a Δ section into a Y section, or vice versa.

Δ-Y transformation. Let Z_A and Z_B be called the adjacents of Z_1; Z_B and Z_C the adjacents of Z_2; Z_C and Z_A *the adjacents of* Z_3. Also let $Z_A + Z_B + Z_C$ be designated as the sum. With this new terminology, Eqs. (7) through (9) state that each

yeasts lose the ability to sporulate upon prolonged cultivation in pure culture.

Imperfect and perfect yeasts Certain haploid yeasts have been shown to be heterothallic, that is, sporulation can occur only if strains of opposite mating type (sometimes indicated by + and − strains) are mixed on sporulation media. If no mating type is known, a yeast is considered to be imperfect. The perfect yeasts are classified in the order Saccharomycetales (syn Endomycetales), and the imperfect yeasts belong in the class Fungi Imperfecti, order Cryptococcales (syn. Torulopsidales) Besides these there are yeasts which produce spores on a short stalk, or sterigma After maturation these so-called ballistospores are shot off by a drop-excretion mechanism. Vegetative reproduction is mainly by budding. These yeasts are members of the order Sporobolomycetales. There is now evidence that they belong to the class Basidiomycetes and the ballistospores are therefore a type of basidiospore (*see* BASIDIOMYCETES; CRYPTOCOCCALES; SACCHAROMYCETALES; SPOROBOLOMYCETALES).

Physiological properties. Some yeasts have the ability to carry out an alcoholic fermentation. Other yeasts lack this property. In addition to the fermentative type of metabolism, fermentative yeasts have as a rule also a respiratory type of metabolism, whereas *nonfermentative* yeasts have only a respiratory, or oxidative, metabolism. Fermentation may be represented by the equation

$$C_6H_{12}O_6 \rightarrow 2C_2H_5OH + 2CO_2$$

and respiration by the equation

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O$$

Both reactions produce energy, with respiration producing by far the most, which is used in part for synthetic reactions, such as assimilation and growth. Part is lost as heat. In addition, small or sometimes large amounts of by-products are formed, including organic acids, esters, aldehydes, glycerol higher alcohols, and others When a fermenting yeast culture is aerated, fermentation is suppressed and respiration increases. This phenomenon is called the reaction of Louis Pasteur.

Fermentation. Yeasts cannot ferment pentose, or 5-carbon sugars (*see* CARBOHYDRATE) In the case of disaccharides, trisaccharides, and polysaccharides, hydrolysis by hydrolytic enzymes, or hydrolases, must precede fermentation There is evidence that such hydrolytic enzymes are located in the peripheral layer of the cell Some hydrolases are actually secreted into the medium as extracellular enzymes Species of yeasts are often differentiated on the basis of the various hydrolases which they possess.

Some simple rules of fermentation have been formulated If a yeast cannot ferment glucose, it cannot ferment any other sugar. If a yeast can ferment glucose, it can also ferment fructose and mannose (although often at different rates), but not always galactose If a yeast ferments maltose, it usually cannot ferment lactose and vice versa.

Respiration and assimilation. Yeasts differ greatly in their ability to respire and to assimilate organic substrates. Depending on the species, yeasts can utilize such compounds as pentoses (D-xylose, D-arabinose, L-arabinose, D-ribose), methylpentoses (L-rhamnose), sugar alcohols (mannitol, sorbitol, dulcitol, erythritol, adonitol), organic acids (lactic, acetic, succinic, citric, gluconic, 2-ketogluconic, 5-ketogluconic, and many other acids), polysaccharides (soluble starch, inulin), and even such compounds as *i*-inositol

The source of nitrogen may be organic or inorganic. All yeasts, with rare exceptions, can use ammonium ion for the synthesis of proteins. Others can utilize nitrate, and still others can use nitrite but not nitrate. Yeasts also can absorb intact amino acids from the medium. There is considerable variation in the ability of yeasts to deaminate individual amino acids. Lysine, as a single nitrogen source, is used by few yeasts Glutamic acid is utilized by nearly all yeasts Sulfur is ordinarily supplied as sulfate, although some yeasts grow better when the sulfur is supplied as cysteine or methionine These mineral salts and concentrations are required by yeasts in a medium: monobasic potassium phosphate (KH_2PO_4), 0.1%, magnesium sulfate ($MgSO_4$) 0.05%; sodium chloride ($NaCl$), 0.01%, and calcium chloride ($CaCl_2$), 0.01% In addition the following trace elements should be added to a synthetic medium: boron, copper, zinc, iron, manganese molybdenum, and iodine.

Special synthetic products are carotenoid pigments produced by species of *Rhodotorula* and *Sporobolomyces* (β-carotene is responsible for a yellow color; torulin and torularhodin are responsible for pink colors). Members of the genus *Cryptococcus* produce a starchlike polysaccharide in acidic media In addition, species of *Cryptococcus*, *Rhodotorula*, and some other yeasts produce capsules, consisting of various polysaccharides

Yeast ecology. Yeasts are not as ubiquitous in nature as are bacteria Many species have highly specific habitats, whereas others are found on a greater variety of substrates in nature.

Soil Soil may be considered as a reservoir of yeasts Active growth in soil only occurs under favorable conditions, such as those found in fruit orchards and meadow lands. However, small numbers of a great variety of yeasts are found in many types of soil. Yeasts belonging to the genera *Lipomyces* and *Schwanniomyces* have been isolated only from soil

Trees Other yeasts have their natural habitat in the exudates or slime fluxes of trees. The yeasts found in exudates of coniferous trees generally differ from those in exudates of deciduous trees Species of *Nadsonia* have only been found in tree exudates; and certain species of *Hansenula*, *Saccharomycodes*, *Pichia*, *Saccharomyces*, and *Endomyces* are specifically associated with certain trees

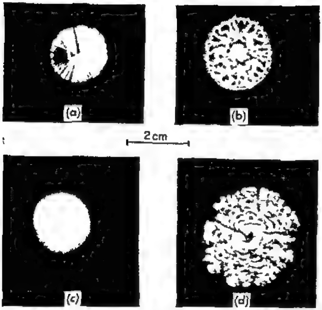


Fig. 3 Giant colonies of *(a) Saccharomyces rouxii; (b) Sporobolomyces pararoseus; (c) Rhodotorula glutinis; (d) Saccharomyces cerevisiae* after 6 weeks of growth on wort gelatin at 20°C.

cultures and as giant colonies. Slant cultures are made by inoculating the yeast as a thin line on the center of an agar slope in a medium which may be malt extract or synthetic. After several weeks of growth the culture may be described with respect to: color—for example cream, pink, yellow, brown, black, white, and intermediate shades; surface—smooth, wrinkled, warty, glossy, dull; texture—pasty, shiny, tough; cross section—flat, raised, hemispherical, extent of spreading; and border—entire, hairy, irregular Giant colonies are usually inoculated on malt gelatin, by a light, pin-point inoculation. Macromorphological features usually develop more characteristically on gelatin than on agar media The description of the growth after about 3 weeks of growth is similar to that used for the description of streak cultures.

Yeast cytology. Yeast cells are surrounded by a wall, which in the case of baker's yeast and many other species consists of polysaccharides such as glucan and mannan, a small amount of chitin, protein, lipids, and minerals The filamentous yeasts have a higher chitin content than the budding yeasts. Some species contain as yet unidentified components With the electron microscope bud scars can be observed in the walls of yeast Successive buds, as many as 23 from one cell, are always formed at different places on the cell surface A cell also contains a birth scar, which differs in appearance from a bud scar. Yeast cells are generally uninucleate. The cytoplasm may contain one or more vacuoles. In addition the cytoplasm may contain lipid globules, volutin (polyphosphate) granules, and submicroscopic particles When yeast cells are treated with iodine, they usually stain deep brown, due to their high glycogen content

Sexual reproduction of yeasts. Yeasts are divided into sporogenous and asporogenous groups, or perfect and imperfect yeasts.

Ascospores. The sexual spores of yeasts are ascospores, which are formed in simple structures, often a vegetative cell. Such asci are called naked asci in the absence of an ascocarp which is a more complex fruiting body found in the higher Ascomycetes. If the vegetative cells are diploid, a cell may transform directly into an ascus after the $2n$ nucleus undergoes a reduction or meiotic division (*see* MEIOSIS). In the case of haploid yeasts, a conjugation or fusion between two cells normally occurs. The fusion of the cytoplasmic contents of the two cells is called plasmogamy. This is followed by fusion of the two haploid nuclei, or karyogamy Meiosis then follows. Each of the haploid nuclei resulting from meiosis gathers cytoplasmic material around it and finally develops into a characteristically shaped ascospore. The spores in such a process are formed by "free cell formation," in contrast to spore delineation by cleavage of the cytoplasm The shape of ascospores may be spherical, oval, kidney- or crescent-shaped, hat-shaped, helmet-shaped, hemispherical, needle-shaped, walnut-shaped, and saturn-shaped

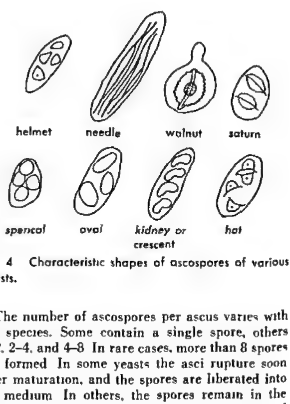


Fig 4 Characteristic shapes of ascospores of various yeasts.

The number of ascospores per ascus varies with the species. Some contain a single spore, others 1–2, 2–4, and 4–8 In rare cases, more than 8 spores are formed In some yeasts the asci rupture soon after maturation, and the spores are liberated into the medium In others, the spores remain in the ascus until they germinate. The heat resistance of ascospores is only slightly greater than that of the vegetative cells The number and shape of the ascospores are important characters in generic descriptions The species of most genera have only one type of ascospores, but sometimes more than one shape is found among the species of a genus, for example, in *Endomycopsis* and *Hansenula*

The ease with which yeasts form ascospores varies enormously with the species. In some, an abundance of ascospores is formed on nearly all media of growth. In others, special sporulation media must be used, and often the percentage of asci is small. Another complication is that certain

Baker's yeast. Baker's yeast is composed almost exclusively of cells of one or more strains of *Saccharomyces cerevisiae*. It is marketed in the form of cakes of compressed fresh yeast cells, whose moisture content is about 70% moisture, or as granules of yeast masses dehydrated to about 7.5% moisture content. Molasses is the principal raw material for the growth of baker's yeast. Depending on cost, availability, and composition, both beet and cane molasses are used in various proportions. Both are by-products of the sugar industry. The molasses is diluted, clarified, and decolorized. It can then be added, as the source of carbohydrate, to the fermentation tanks. These are vigorously aerated so as to suppress fermentation and alcohol production; in the presence of oxygen the yeast metabolism is chiefly oxidative and the crop per unit amount of raw material used may be increased by a factor of 20. Since the oxidative metabolism yields also much heat, the contents of the tanks are cooled by water flowing through a series of coils. The temperature is kept at 26–30°C, because higher temperatures influence the quality of the yeast unfavorably. The pH is maintained at 3.8–4.5. At higher pH values the danger of development of bacterial infection increases markedly.

Nitrogen is supplied in the form of ammonium sulfate and ammonium hydroxide; phosphate as dibasic sodium phosphate, Na_2HPO_4, and magnesium as magnesium sulfate, $MgSO_4$. Potassium is usually present in sufficient amounts in molasses. The sugar is added slowly and continuously so that its total concentration remains very low, thus reducing the amount of alcohol formed. After all nutrients have been added, aeration and carbohydrate addition are continued in order to ripen the yeast. During this period young buds mature, the yeast increases its content of reserve carbohydrates and improves as far as storage stability is concerned. Foaming of the culture liquid, or wort, during the period of aeration is controlled by the continuous addition of various antifoam compounds.

Compressed cake manufacture. After completion of growth the wort is centrifuged and concentrated about tenfold. The resulting yeast cream, containing about 4 lb of yeast per gal, is washed and centrifuged again after which it is pressed in a filter press. The yeast is scraped from the presses and placed in a mixer where a small amount of vegetable oils is added to facilitate cutting and to improve the appearance. The yeast is extruded as a solid bar from a special machine and cut into blocks of specified weight. The blocks are then wrapped and refrigerated to about 32°F. Starting with the cream stage, the yeast is artificially cooled for greater stability and to reduce contamination. Packing is also done at reduced temperatures.

Active dried yeast manufacture. For the manufacture of active dried yeast for baking purposes specially selected strains are used. The yeast is treated in the fermentors in such a way that its reserve carbohydrate content, in the form of trehalose and glycogen, becomes high. The compressed yeast is usually extruded from a machine into thin strands which are dried on trays in an air flow of carefully controlled temperature and humidity. When dried to about 7.5% moisture content, the yeast is packed in moistureproof containers, usually in the absence of oxygen, and refrigerated. Some plants use rotary driers, which produce smooth round pellets, rather than short spaghettilike pieces of porous appearance. Dried yeast is stable for many months under refrigeration. Rehydration of dried yeast must be done in water of about 43°C. Lower temperatures cause extensive leaching of the cell contents and a loss in baking quality. Standardized baking tests are conducted to evaluate the quality of the product.

Food and fodder yeasts. These are made for human and animal consumption and represent valuable additional sources of proteins, vitamins of the B-complex, and minerals. They are produced from a variety of raw materials and by widely different processes. Some, known as secondary yeasts, represent a by-product of the brewery and distillery industries; others, the primary yeasts, are grown specifically as food sources from cane or beet molasses, hydrolyzed starch, spent sulfite liquor of the paper industry, acid-hydrolyzed wood, and agricultural wastes. In some cases secondary yeast, usually brewer's yeast, is prepared by letting the crop harvested from the beer vats grow for one or more generations in molasses, with aeration; this is done to eliminate the bitter taste of the yeast due to the hops used in the beer wort.

Raw materials. The choice of raw materials for yeast manufacture usually depends on their cost and availability. Excellent sources of carbohydrates are beet molasses, blackstrap molasses of the cane sugar industry, and hydrol, the residual syrup from the corn-sugar industry. An inexpensive substrate is spent sulfite liquor, which is a waste product of the sulfite-pulping process in the paper industry. About 400 lb of carbon compounds utilizable by yeast are contained in the waste liquor from each ton of pulp. The carbon compounds consist mainly of hexoses, pentoses, and acetic acid, but their proportions vary greatly with the type of wood used. Since sulfur dioxide, SO_2, is toxic to yeast, steam stripping or a similar process is applied to remove as much SO_2 as is possible. Neutralization by lime and a settling process further reduce the SO_2 content.

Another source of raw materials is acid-hydrolyzed wood, as obtained in the Scholler or the Bergius process. The production of wood sugar by these methods is fairly costly so that manufacture of food yeast from this source is less economical than that from sulfite waste liquor. Agricultural wastes form another source of carbohydrates for the growth of yeast. These include milk whey, fruit waste, citrus peel hydrolyzates, cull potatoes, and others. Their seasonal supply is a disadvantage.

Yeast strain. The choice of yeast depends on the substrate and growing conditions. The *Saccharomyces* yeasts are only suitable with molasses and

Insects. Since insects, such as *Drosophila*, frequently use sap exudates for breeding and sometimes for feeding purposes, yeasts can be found in appreciable numbers in the intestinal tract of such insects. Bark beetles, which live in the cambium layer of coniferous as well as deciduous trees, and the wood-boring ambrosia beetles also have been shown to have a specific yeast flora associated with them. Many species of yeast have been found in nature only in association with certain insects. Yeasts are rapidly digested by insects, and in this way they form an important part of their diet. The nectar of flowers is another habitat for yeasts. Especially those flowers which are much frequented by bees may contain large numbers of yeast. The intestinal tract of warm-blooded animals is also a habitat for yeasts. In some cases the yeast is so dependent on its host that it has lost the ability to grow at room temperature and in the common media used for propagation of yeasts. An example is the monospecific genus *Saccharomycopsis*, which has been found only in the intestinal tract of rabbits. *Saccharomycopsis guttulata* only grows at about 37°C and in very complex media.

Pathogens. Whereas most yeasts associated with warm-blooded animals are nonpathogenic, a number of at least potentially pathogenic yeasts are known. With the expanding use of antibiotics, the natural balance of microbes in the gastrointestinal tract of mammals appears to be disturbed, and yeast infections are on the increase. The most common locations of infections by yeasts are the skin, especially the mucosa, and the respiratory tract, but occasionally systemic infections occur (*see* MYCOLOGY, MEDICAL).

Candida albicans is the most common cause of yeast infections in man. In medical literature, it is often referred to as *Monilia*. The basic cause of pathogenicity of *C. albicans* is not clear, since many normal individuals, while carriers of this yeast, experience little if any ill effects. Diseases caused by *C. albicans* include thrush in the oral cavity of infants, infections of the skin, nails, bronchi, lungs, vagina, and the intestinal tract. Other species of *Candida*, such as *C. tropicalis*, *C. parapsilosis*, *C. guilliermondii*, are more rarely associated with yeast infections. In general, the diseases in which species of *Candida* are involved are called candidiasis or moniliasis (*see* CANDIDIASIS). *Cryptococcus neoformans* (syn. *Torulopsis histolytica*, *Cryptococcus hominis*) causes a disease called cryptococcosis or torulosis (*see* CRYPTOCOCCOSIS). It is relatively rare, but often fatal. In man, the yeast develops in the central nervous system and induces a chronic and often fatal meningitis. Sporadic cases have been reported from all [illegible] but no epidemic has ever been reported, [illegible] contagious [illegible] nature, but it has been found rather regularly in pigeon droppings and rarely in soil. Another yeast which may occasionally become pathogenic is *Torulopsis glabrata*. It has been found associated with diseases similar to those caused by *Candida albicans*, and occasionally it may cause a systemic infection. *Pityrosporum ovale* (bottle bacillus) can frequently be found on the skin, especially the scalp, of persons suffering of seborrheic dermatitis or pityriasis (dandruff). Although some believe that this yeast is responsible for these diseases, most of the evidence points to the fact that *P. ovale* is a harmless saprophyte of man and that it occurs on the scalp because of its requirement of oils or fats for growth.

Some mycotic infections by true fungi are characterized by yeastlike cells in the infected tissues (yeast phase), whereas in culture the causative organisms produce hyphae and look like molds (mycelial saprophytic phase). This is called the phenomenon of dimorphism. An example is *Blastomyces dermatitidis*, the causal agent of North American blastomycosis (*see* BLASTOMYCOSIS).

Plant pathogenic yeasts are limited to a few species. These are *Nematospora coryli*, a yeast with needle-shaped ascospores, and a yeastlike fungus *Ashbya gossypii* producing similar ascospores. The diseases which result from infection by *N. coryli* are yeast-spot in Lima beans and soybeans; stigmatomycosis of the fruits of tomatoes, citrus fruits, and others, and discoloration of cotton bolls. The last plant is also the principal host for *A. gossypii*. The yeast is transmitted by plant bugs and other biting insects (*see* PLANT DISEASE).

Yeasts in food spoilage. Yeasts are important as spoilage organisms of a great variety of foods. Specific types occur, depending primarily on the composition of the food. Sugar-tolerant, or osmophilic yeasts occur on dried fruits, syrups, honey, and the like. Examples are *Saccharomyces rouxii*, *Schizosaccharomyces octosporus*, and the yeastlike fungus *Eremascus albus*. Species of *Debaryomyces* are usually highly salt-tolerant and occur in brines and on processed meats, such as bacon, hams, and sausage. Dairy products may be spoiled by lactose-fermenting yeasts. Some species of *Torulopsis* and of *Zygosaccharomyces* (haploid forms of *Saccharomyces*) may spoil salad dressing and similar products, because of their great tolerance to vinegar. *Endomycopsis fibuliger* forms starch-splitting enzymes and can cause spoilage of cereal products and grains. *See* FOOD MICROBIOLOGY.

[E.M.M.; H.J.P.]

Bibliography: A. H. Cook (ed.), *The Chemistry and Biology of Yeasts*, 1958; J. Lodder and N. J. W. Kreger-Van Rij, *The Yeasts—a Taxonomic Study*, 1952.

Yeast, industrial

The yeast used for the production of fermented beverages or other fermented foods, for baking purposes, for the production of vitamins, proteins (food and fodder yeast), alcohol, glycerol, and enzymes. For each of these processes definite species of yeast are employed, and in many cases special strains are selected for optimal results.

experience with some other member of the group.

"Urban" yellow fever, transmitted in a cycle of man to *Aedes aegypti* to man, has been virtually eliminated by killing the mosquito vector. However, workers in jungle areas where no *Aedes aegypti* exist can become infected with "jungle yellow fever," which is primarily a disease of monkeys, transmitted by mosquitoes of the hemogogus group in South America, and by other mosquito species in other areas.

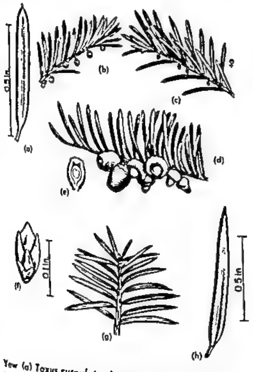
Prophylactic control is readily achieved by administration of an attenuated live vaccine, in the form of the 17D strain, derived by serial passage in tissue culture. *See* ANTIBODY; CULTURE, EMBRYONATED EGG; CULTURE, TISSUE; SERUM.

[J.L.M.]

Bibliography: T. M. Rivers and F. L. Horsfall, Jr (eds.), *Viral and Rickettsial Infections of Man*, 3d ed., 1959

Yew

A genus of evergreen trees and shrubs, *Taxus*, with a fruit containing a single seed surrounded by a scarlet, fleshy, cuplike envelope (aril). The leaves are flat and acicular (needle-shaped), green below, with stalks extending downward on the



Yew (a) *Taxus cuspidata*, Japanese yew, leaf (b) Male branchlet, natural size, showing staminate flower buds (c) Female branchlet, natural size, showing ovule-bearing flowers. (d) Female branchlet with various stages of development of fleshy aril around seed. (e) Lengthwise section of berry showing position of seed in aril (f) *T. baccata*, English yew, bud, showing obtuse, scarcely keeled scales (g) Branchlet, natural size (h) Leaf (From A H Graves, *Illustrated Guide to Trees and Shrubs*, rev ed., Harper, 1956)

stem The only native American species of commercial importance is the Pacific yew, *T. brevifolia*, a medium sized tree of the Pacific Coast and northern Rocky Mountain regions Although it is not a common tree, its wood is sometimes used for poles, paddles, bows, and small cabinet work.

The English yew, *T. baccata*, native in Europe, North Africa, and northern Asia, and the Japanese yew, *T. cuspidata*, are much cultivated in the United States as evergreen ornamentals. Both are small trees when mature, but often in cultivation they are pruned to lower dimensions. In the English yew the leaves taper gradually to a point and are glossy, whereas in the Japanese yew the leaves are abruptly pointed and of duller appearance. There are many cultivated forms of these two species.

The Canada yew, *T canadensis*, also known as ground hemlock, is a low, straggling shrub of the forests of the northeastern quarter of the United States. It is distinguishable from a true hemlock by the absence of white lines on the underside of the leaves. *See* FOREST AND FORESTRY; TREE.

[A.H G.]

Yolk sac

A saclike extraembryonic membrane present in and enclosing a relatively large yolk mass in bony fishes, elasmobranchs (sharks), and all reptiles, birds, and primitive mammals (Prototheria); or containing a small, variable amount of fluid yolk in higher mammals Typically, the yolk sac is attached to the lower, or ventral, part of the body of the embryo *See* FETAL MEMBRANE.

The yolk sac wall of amniotes (reptiles, birds, and mammals) is composed of an inner layer of cells, continuous with and actually constituting part of the endodermal (hypoblast) germ layer. The outer, or surface, layer of the yolk sac is composed of mesodermal (middle germ layer) cells continuous with the mesoderm of the embryo proper The two layers are collectively termed the splanchnopleure The cavity of the yolk sac is continuous with that of the primitive gut, or digestive tract, of the embryo Even after completion of the process of constriction of the embryonic membranes below the embryo to form the umbilical cord in mammals, or its homolog in reptiles and birds, the yolk sac remains attached to the embryo by a narrow tube, the yolk stalk

The primary function of the yolk sac in reptiles, birds, primitive mammals and those fishes and sharks that have it, is nutritive The wall of the yolk sac is laced with blood vessels and thrown into deep folds or villous projections Nutrients in the yolk are absorbed by the blood vessels and transported to the embryo. Ample evidence attests to an important digestive function of the yolk sac in converting yolk nutrients into transportable forms with the aid of enzymes. *See* GERM LAYERS

[N T S]

Bibliography L. B Arey, *Developmental Anatomy*, 6th ed., 1954, J Needham, *Biochemistry and Morphogenesis*, 1942.

hydrol as the substrate since they do not utilize pentose sugars well. *Candida utilis* (*Torulopsis utilis*, Torula-yeast) can utilize a much greater array of compounds than *S. cerevisiae* and it respires pentoses (five-carbon sugars) as well as hexoses (six-carbon sugars). It is the principal species grown on wood hydrolyzate and on sulfite liquor. Two varieties of this yeast are used. *T. utilis* var. *major* has large cells and can be harvested more easily. *T. utilis* var. *thermophilus* grows well at 36–39°C and is useful where temperature control of the fermentors is difficult, or in the tropics. The yeastlike organism *Oospora lactis* (*Oidium lactis, Geotrichum candidum*) has also been used, but its digestibility has been questioned.

Production process. Regardless of the strain, the yeast is grown in large fermentors under highly aerated conditions while carbohydrates and sources of nitrogen, phosphorus, and potassium are added at a rate which parallels the growth rate of the yeast. Some processes operate on a continuous basis, and the rate of substrate addition equals the rate of withdrawal of yeast-containing effluent. At least six different commercial processes are in use. The yeast suspension is concentrated by centrifugation and usually dried on steam heated rotary drum driers. The dried yeast is valuable in protein deficient areas of the world for human consumption and is used in animal feed mixtures, especially poultry feed.

Vitamin production by yeasts. Many microorganisms, including the yeasts, can synthesize vitamins, especially of the B-complex (*see* VITAMIN). But there is a marked variation in this respect, even among different strains of the same species. Some yeasts can grow only if supplied with one or more vitamins or their precursors, others can grow in a vitamin-free medium, and then synthesize those they need for their own functioning.

Yeasts may also absorb vitamins, such as thiamin, nicotinic acid, and biotin, from the medium and store them in concentrations far above normal; thus they can yield enriched products of predetermined vitamin content. Baker's and brewer's yeast, as well as *Candida utilis*, are good sources of thiamin; riboflavin; pyridoxin; biotin; pantothenic, nicotinic, folic, and *p*-aminobenzoic acids; choline; and inositol. Some yeasts produce exceptionally large amounts of a particular vitamin, *Eremothecium ashbyi* and *Ashbya gossypii*, for example, can synthesize more than 2 mg of riboflavin per ml of medium, and this biologically produced riboflavin can readily compete on the market with the chemically synthesized product (*see* RIBOFLAVIN). Strains of *Saccharomyces* can produce ergosterol so that it may amount to as much as 7–10% of the dry weight of the yeast; this ergosterol can be transformed into vitamin D_2 (calciferol), a valuable feed supplement for mammals, by ultraviolet irradiation.

Glycerol production by yeast. Normally glycerol is produced as a minor by-product in the alcoholic fermentation of sugar by yeast. But in nonaerated cultures, supplied with calcium sulfite and maintained at a pH of 6.7–7.0, yeast can convert 27% of the fermentable sugar to glycerol. Another process giving good yields of glycerol is the alkaline fermentation carried out in the presence of 30% sodium carbonate, in terms of fermentable sugar. Selection of suitable strains is imperative.

Yeast as a source of enzymes. Invertase is the principal commercial enzyme obtained from baker's or brewer's yeast. It hydrolyzes sucrose to a mixture of glucose and fructose which is called invert sugar. Invertase is an intracellular enzyme and is obtained from yeast by treatment with toluene. This causes autolysis, or self-digestion, and invertase is released in soluble form. Purified invertase is used in the manufacture of artificial honey and invert sugar. It is also used to make chocolate-coated soft-cream-center bonbons. Hard centers are prepared with sucrose plus invertase; and after they have been coated with chocolate, the enzyme gradually transforms the sugar into a syrupy consistency owing to the high solubility of the invert sugar. Lactase, which splits lactose into glucose and galactose, is prepared from *Saccharomyces fragilis*. It is used to prevent crystallization of lactose in ice cream and in concentrated milk and whey. Some strains of *S. fragilis* produce an extracellular pectic enzyme, polygalacturonase, of high purity. *See* ENZYME (INDUSTRIAL PRODUCTION); INDUSTRIAL MICROBIOLOGY; YEAST.

[F.M.M ; H.J.P.]

Bibliography: L. A. Underkofler and R. J. Hickey (eds.), *Industrial fermentations*, 2 vols., 1953–1954.

Yellow fever

An acute, febrile, mosquito-borne viral disease characterized in severe cases by jaundice, albuminuria, and hemorrhage. Inapparent infections also occur.

The agent is an arbor group B virus, about 25 millimicrons in diameter. The arbor group viruses are arthropod-borne viruses. The virus multiplies in many animals in mosquitoes, chick embryos, and tissue cultures from chick or mouse embryos. Freshly isolated strains are pantropic, that is, they invade tissues derived from all three embryonal layers. Serial animal or culture passages can modify these tropisms. *See* ANIMAL VIRUS, ARBOR VIRAL ENCEPHALITIDES.

The virus enters the body through a mosquito bite and multiplies in lymph nodes, circulates in the blood, and localizes in the liver, spleen, kidney, bone marrow, and lymph glands. The severity of the disease and the major signs and symptoms which appear depend upon where the virus localizes and how much cell destruction occurs.

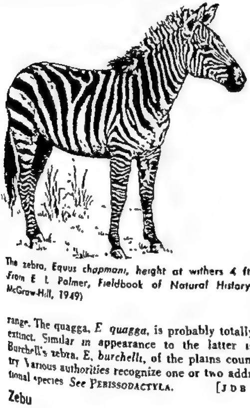
Diagnosis is made by isolation of the virus from serum obtained from a patient as early as possible in the disease and inoculated intracerebrally in mice, or by the rise in serum antibody. Antibody response may be broad, and may include other group B arbor viruses, if there has been previous

Z Zebra to Zythiaceae

Zebra

Species of striped animals of the family Equidae, all found in Africa. They are somewhat more like asses than true horses in general anatomy but are easily distinguished by the prominent stripes. The zebras are gregarious animals typical of the grassy plains but are also found in the mountains south of the Sahara Desert. They commonly travel in small herds, frequently in the company of gnus. Zebras are heavily preyed upon by lions and are hunted by natives for food and for hides. They have been domesticated to a limited degree.

The common, or mountain zebra, *Equus chapmani* (zebra), formerly occurred in the mountains of South Africa but is now limited to a few protected areas in the eastern part of its original

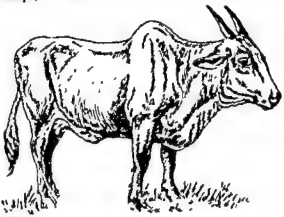


The zebra, *Equus chapmani*, height at withers 4 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

range. The quagga, *E. quagga*, is probably totally extinct. Similar in appearance to the latter is Burchell's zebra, *E. burchelli*, of the plains country. Various authorities recognize one or two additional species. See PERISSODACTYLA. [J.D.B.]

Zebu

Better known in the United States as the Brahma, *Bos indicus*, also called East Indian ox (the sacred cow of Indian Hindus), is found in southern Asia and Africa. African strains have large horns, in contrast to the Indian breed which has short horns or is hornless. Zebus are thick-chested, rangy animals, the bulky appearance of their foreparts emphasized by the prominent hump, sometimes two humps, between their shoulders. In India they are



The zebu, *Bos indicus*. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

widely used as work animals. In the United States, southern cattle breeders have successfully mated Brahma bulls with domestic females to produce an animal more resistant to ticks and heat than heretofore possible. See ARTIODACTYLA. [J.D.B.]

Zeeman effect

A splitting of spectral lines when the light source being studied is placed in a magnetic field. First discovered by P. Zeeman in 1896, the Zeeman effect furnishes information of prime importance in the analysis of spectra. Each kind of spectral term has its characteristic mode of splitting, and the types of terms are most definitely identified by this property. Furthermore, the effect allows an evaluation of the ratio of charge to mass of the electron and an evaluation of its precise magnetic moment.

Normal Zeeman effect. This is a splitting into two or three lines, depending on the direction of observation, as shown in Fig. 1. The light of these components is polarized in ways indicated in the figure. The normal effect is observed for all lines belonging to singlet systems, those for which the spin quantum number $S = 0$. The change of frequency $\Delta\nu_n$ of the shifted components can be evaluated on classical electromagnetic principles as follows. Assume that the electron of charge e revolves in a circular orbit of radius r and circular frequency ω radians per second (Fig. 2). If a magnetic field H is applied perpendicular to the plane of the orbit, the electron will be speeded up or slowed

Young's modulus

A constant designated E, the ratio of stress to corresponding strain when the material behaves elastically. Young's modulus is represented by the slope $E = \Delta S/\Delta \epsilon$ of the initial straight segment of the stress-strain diagram (*see* STRESS AND STRAIN). More correctly, E is a measure of stiffness, having the same units as stress, pounds per square inch. When stress and strain are not directly proportional, E may be represented as the slope of the tangent or the slope of the secant connecting two points on the stress-strain curve. The modulus is then designated as tangent modulus or secant modulus at stated values of stress. The modulus of elasticity applying specifically to tension is called Young's modulus. Many materials have the same value in compression. *See* HOOKE'S LAW. [W.J KR.]

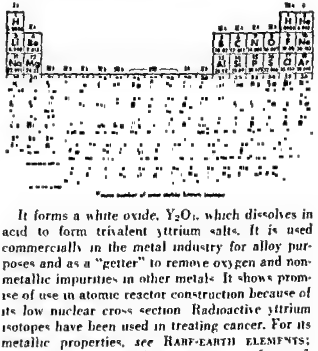
Ytterbium

Element number 70, ytterbium, Yb, is a metallic element belonging to the rare-earth group. Its atomic weight is 173 04 and the naturally occurring stable isotopes are Yb^{168} 0 135%, Yb^{170} 3.03%, Yb^{171} 14.31%, Yb^{172} 21.82%, Yb^{173} 16.13%, Yb^{174} 31.84%, Yb^{176} 12.73% Its compounds were first separated by J. C. G. Marignac in 1878 Later, G. Urbain in 1907 showed that Marignac's ytterbium was composed of two elements, lutetium and what is now known as ytterbium It has also sometimes been called aldebaranium. The common oxide, Yb_2O_3, is colorless and dissolves readily in acids to form colorless solutions of trivalent salts which are also paramagnetic.

Ytterbium also forms a series of divalent compounds. The divalent salts are soluble in water but react very slowly with water to liberate hydrogen Ytterbium can be separated from the other rare earths by taking advantage of its divalent state, or it can be removed from other rare earths by extracting it with sodium amalgam. The metal is best prepared by distillation (*see* EUROPIUM; SAMARIUM). It is a silvery soft metal which corrodes slowly in air and resembles the calcium-strontium-barium series more than the rare earth series, since the metal probably has two electrons in the conduction bands instead of three. *See* RARE-EARTH ELEMENTS. [F.H.SP.]

Yttrium

Element number 39, yttrium, Y, closely resembles the rare-earth elements. It has atomic weight 88 92, and the stable isotope Y^{89} makes up 100% of the naturally occurring element. It is always found associated with the rare earths and is frequently classified as one of them J. Gadolin discovered the element in 1794, and it was obtained in high purity by C. G. Mosander in 1843



It forms a white oxide, Y_2O_3, which dissolves in acid to form trivalent yttrium salts. It is used commercially in the metal industry for alloy purposes and as a "getter" to remove oxygen and nonmetallic impurities in other metals It shows promise of use in atomic reactor construction because of its low nuclear cross section Radioactive yttrium isotopes have been used in treating cancer. For its metallic properties, *see* RARE-EARTH ELEMENTS; *see also* LANTHANUM; SCANDIUM [F.H.SP.]

and S are finite, the effective magnetic moment may be written

$$\mu_J = g(eh/4\pi mc)J = g\mu_0 J$$

where J measures the total angular momentum (in LS coupling, the resultant of L and S), and μ_0 is the Bohr magneton, $eh/4\pi mc$ (see PARAMAGNETISM). Theory gives values of g, the Landé g factor, which are characteristic of the type of spectral term. In LS coupling, the value is given by

$$g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}$$

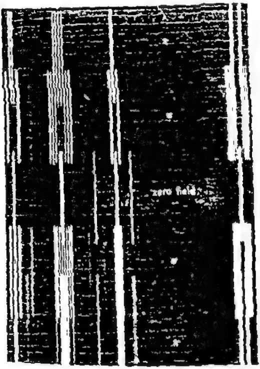


Fig 4 Zeeman effect of the rhodium spectrum in the wavelength range 3479–3462 A. Field strengths, 70,-000 oersteds (lower exposure) and 90,500 (upper exposure). (G R. Harrison and F Bitter, Massachusetts Institute of Technology)

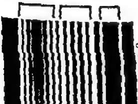


Fig 5 Zeeman effect of three adjacent rotational lines of the hydrogen molecule (Johns Hopkins University)

For a classical electron orbit $g = 1$, which yields the normal Zeeman effect. When the spin S is present, however, the changes in energy produced by the magnetic field, which are proportional to μ_J, are just g times as great, and this fact is responsible for the anomalous Zeeman effect. It should also be mentioned that both theory and experiment now show that the g factor for the electron is not exactly 2, but 2.00229.

Quadratic Zeeman effect. The quadratic effect, which depends on the square of the field strength, is of two kinds. The first results from the second-order terms that were neglected in the preceding derivation, and the second, from the diamagnetic reaction of the electron when revolving in large orbits.

Inverse Zeeman effect. This is the Zeeman effect of absorption lines. It is closely related to the Faraday effect, the rotation of plane-polarized light by matter situated in a magnetic field. *See* FARADAY EFFECT, *see also* ATOMIC STRUCTURE AND SPECTRA. [F.A.J.]

Zeeman effect in molecules. This effect is, in general, so small as to be unobservable, even for molecules which have a permanent magnetic moment. Each level with a total angular momentum J splits into $2J + 1$ components, as in the case of atoms. The component of the magnetic moment along the direction of the external field is small, however, because the rotation of the molecule, which carries the magnetic moment along with it, causes the principal part of the magnetic moment to average out to zero. The consequence is that the magnetic levels have an extremely narrow spacing except for those cases where the molecule has either very little rotation or none at all. An exception occurs for some light molecules where the magnetic moment is coupled so lightly to the frame of the molecule that it can orient itself freely in the magnetic field just as for atoms (see Fig. 5).

Zeeman effect in crystals. A clear Zeeman effect can also be observed in many crystals with sharp spectrum lines in absorption or fluorescence. Such crystals are found particularly among the salts of the rare earths. In these cases the internal electric field in the crystal splits and shifts the level of the free ion. When the number of electrons is even and the crystal symmetry low, this electric splitting is complete. No degeneracy remains, and there can be no further splitting by a magnetic field. If the number of electrons is odd, or if for an even number there is high crystal symmetry, the levels occur in degenerate pairs which are split by a magnetic field. Each line is then split into four components

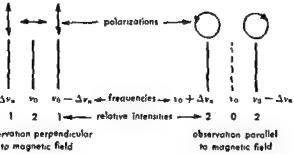


Fig. 1. Triplet observed in the normal Zeeman effect

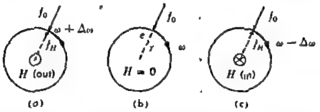


Fig 2. (a, b, c) Effect of a magnetic field H applied perpendicular to a circular electron orbit.

down because of the changing flux through its orbit, just as in the electron accelerator known as the betatron

Denoting the centripetal force holding the electron in its orbit before application of the field by f_0 (Fig. 2*b*), and the additional force due to the motion of the electron across the field by f_H (Fig 2*a* and *c*), one has

$$f_o = m\omega^2 r \quad \text{and} \quad f_H = He(\omega \pm \Delta\omega)r$$

In Fig. 2*a*, where the two forces are in the same direction, one may equate their sum to the centripetal force on an electron of frequency $\omega + \Delta\omega$, obtaining

$$f_o + f_H = m\omega^2 r + He(\omega + \Delta\omega)r = m(\omega + \Delta\omega)^2 r$$

Solution of this equation for $\Delta\omega$, under the assumption that it is small compared to ω itself, yields

$$\Delta\omega = eH/2m \quad \text{and} \quad \Delta\nu_n = eH/4\pi m$$

the latter expression following from the fact that $\nu = \omega/2\pi$.

This relation, although derived for a special case, is generally valid for any system of particles having a particular value of e/m and moving under the action of a central force (*see* LARMOR PRECESSION). On substitution of the ratio of charge to mass of an electron, one obtains

$$\Delta\nu_n = 1.3996 \times 10^6 H \text{ sec}^{-1}$$

Conversely, from the observed spectroscopic splitting $\Delta\nu_n$, and measurement of the field strength, the value of e/m for the electron has been evaluated as 1.7572 ± 0.0007 emu/g. This is in good agreement with the figure determined by other methods.

Anomalous Zeeman effect. This effect is a more complicated type of line splitting, so named because it did not agree with the predictions of classical theory. It occurs for any spectral line arising from a combination of terms of multiplicity greater than one. As examples, Fig. 3 gives diagrams of the theoretical patterns for the yellow lines of sodium, belonging to a doublet system, while Fig. 4 shows some actual patterns observed for doublets and quartets in rhodium.

Since multiplicity in spectral lines is caused by the presence of a resultant spin vector S of the electrons, the anomalous effect must be attributed to a nonclassical magnetic behavior of the electron spin. While classical theory associates with the vector L of the orbital angular momentum a magnetic moment

$$\mu_L = (eh/4\pi mc)L$$

it is necessary, in explaining the anomalous Zeeman effect, that the magnetic moment corresponding to S be

$$\mu_S = (eh/2\pi mc)S$$

Thus the spin generates twice as much magnetic moment, relative to its angular momentum, as does the orbital motion In an atom for which both L

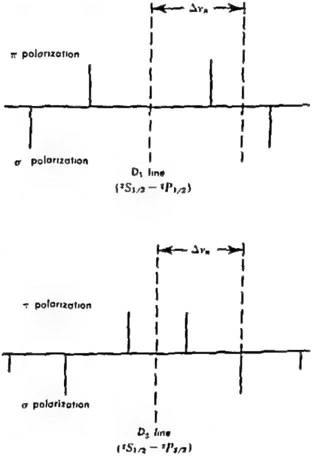


Fig 3. Anomalous Zeeman effect of the sodium lines $\Delta\nu_n$ denotes the normal Zeeman splitting, while π and σ refer to polarizations like those of the central and outer components in the normal effect illustrated in Fig 1. The heights of the lines indicate their relative intensities

acid When placed in the flame they appear to boil as they fuse to a blebby glass.

In gross chemical composition the zeolites resemble the feldspars in that they are aluminosilicates containing sodium, calcium, and potassium. Like the feldspars, they are tectosilicates built on a three-dimensional framework of linked SiO_4 tetrahedra in which aluminum takes the place of some of the silicon The excess negative charge resulting from the substitution of Al for Si is compensated for by the introduction in the structure of Ca, Na, and K *See* SILICATE MINERALS

The crystallography of the zeolites differs greatly from one to another, since they crystallize in the isometric, orthorhombic, monoclinic, and hexagonal systems Nevertheless, all have similar open structures and contain channelways in which water is readily housed This water differs from that contained in most minerals, for as the zeolite is heated, the water is given off continuously, leaving the structure intact After nearly complete dehydration, the channels may again be filled with water or other materials such as ammonia, iodine, or alcohol. This process is selective depending on the relative sizes of the molecules and the channelways In a given zeolite certain large molecules will be excluded whereas smaller ones will be admitted; hence zeolites are used as molecular sieves Crystalline substances resembling natural zeolites are manufactured for this purpose *See* MOLECULAR SIEVE.

Zeolites have another interesting property based on the structure, that of base exchange or cation exchange Through this agency zeolites or synthetic substances with zeolitic structure are used

Some minerals in the zeolite family

Natrolite group	
Natrolite	$Na_2[Al_2Si_3O_{10}]\ 2H_2O$
Mesolite	$Na_2Ca_2[Al_2Si_3O_{10}]_3\ 8H_2O$
Thomsonite	$NaCa_2[Al_2(Al\ Si)Si_2O_{10}]_2\ 5H_2O$
Scolecite	$Ca[Al_2Si_3O_{10}]\ 3H_2O$
Laumontite-gismondite group	
Gismondite	$Ca[Al_2Si_2O_8]\ 4H_2O$
Laumontite	$Ca[AlSi_2O_6]_2\ 4H_2O$
Mordenite	$(Ca,K_2,Na_2)[AlSi_5O_{12}]_2\ 7H_2O$
Heulandite-stilbite group	
Heulandite	$Ca[Al_2Si_7O_{18}]\ 6H_2O$
Stilbite	$Ca[Al_2Si_7O_{18}]\ 7H_2O$
Epistilbite	$Ca[Al_2Si_6O_{16}]\ 5H_2O$
Brewsterite	$(Sr,Ba,Ca)[Al_2Si_6O_{16}]\ 5H_2O$
Phillipsite group	
Phillipsite	$KCa[Al_3Si_5O_{16}]\ 6H_2O$
Harmotome	$Ba[Al_2Si_6O_{16}]\ 6H_2O$
Chabazite group	
Gmelinite	$(Na_2,Ca)[Al_2Si_4O_{12}]\ 6H_2O$
Chabazite	$(Ca,Na_2)[Al_2Si_4O_{12}]\ 6H_2O$
Analcime*	$Na[AlSi_2O_6]\ H_2O$

* Chemically analcime is closely related to leucite and is sometimes considered a member of the feldspathoid group

in water softeners. This is accompl[illegible]
hard water t
a sodium ze
change plac

making the water soft. When the zeolite has had all the sodium removed, it can be regenerated by passing a sodium chloride brine through the container and forcing the exchange in the opposite direction. *See* WATER SOFTENING. [C S HU.]

Zero

In mathematics, the concept zero is used in two ways: as a number and as a value of a variable. The positional system of number notation, developed first by the Babylonians with the basis 60, and later by the Hindus and the Chinese with the base 10, required for greater clarity a special marker of the empty, nonoccupied position. As such, the symbol for zero was introduced by the Babylonians (about 500 B C.), and a millennium later in the Indian system of notation, which came to the West through the Arabs as the arabic number system.

The zero as a number, however, is a new concept, introduced by the Hindus and Chinese about the same time (sixth century). Brahmagupta (born A D 598) remarked that the number 0 has special properties: $a \pm 0 = a$, and $a \cdot 0 = 0$, where a may be any number (integer) Using zero as a denominator, he stated $a/0 \pm b = a/0$, where thus $a/0$ appears as a number of a new kind, which cannot be changed by addition or subtraction. (In the modern way of thinking, division by zero is therefore not a "permissible" operation) In the acceptance of zero as a number, the Orient is centuries ahead of the Occident. Not until Leonardo Fibonacci (*Liber abaci*, 1228) was the zero treated It is of interest that the negative numbers appear earlier in the history of mathematics than does the number zero, they were already known to the Chinese mathematician Liu Hui in the third century A D

In a modern way zero can be called the identity element of the infinite Abelian additive group of integers If in an integral domain (a fortiori in a field) a product is equal to zero then at least one factor of the product is zero

In the second concept zero is the value of a variable for which a function is equal to zero For example, "A polynomial of degree n has n zeros," or "The Riemann zeta function $\zeta(s)$ has all its complex zeros in the strip $0 <$ real part $s < 1$" *See* NUMBER THEORY [H R.]

Zinc

Chemical element number 30, zinc, Zn, is a malleable, ductile gray metal with atomic weight 65 38. Because of chemical similarities among zinc, cadmium and mercury, these three metals are classed together in group IIb of the periodic table of elements Zinc became known in Europe during the Middle Ages, although it had been known much earlier in Asia Paracelsus, in the sixteenth century, is said to have been the first European to recognize zinc as a distinct metallic element and to call it "zinckum."

Thirteen isotopes of zinc are known, of which five are stable, having atomic masses of 64, 66, 67,

(Fig. 6). For cubic crystal symmetry and when the angular momentum is entirely due to the electron spin, a splitting into more than four components may occur.

Nuclear Zeeman effect. The magnetic moment of the nucleus causes a Zeeman splitting in atomic spectra which is of an order of magnitude 1000

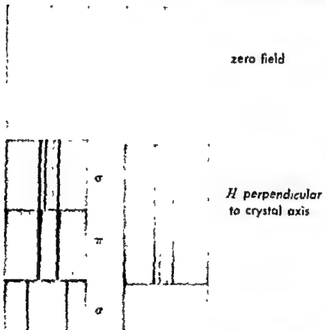


Fig 6. Zeeman effect of an absorption line of neodymium chloride ($NdCl_3$), left with, right without polarization. Notice the splitting into four components. The splitting is quite different depending on whether the trigonal crystal axis is parallel or perpendicular to the magnetic field H = 35,000 oersteds (Johns Hopkins University)

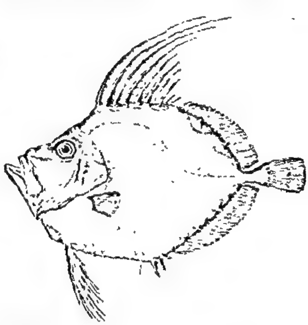
smaller than the ordinary Zeeman effect. This Zeeman effect of the hyperfine structure is usually modified by a nuclear Paschen-Back effect. See PASCHEN-BACK EFFECT.

A strong magnetic field may actually modify the intensity and selection rules so that usually absent lines may appear. For example the *J*-selection rule is no longer valid in a magnetic field. [C H D]

Bibliography: F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3d ed., 1957; J. H. Van Vleck, *The Theory of Electric and Magnetic Susceptibilities*, 1932.

Zeiformes

A small order of teleost fishes, of which the dories are structurally intermediate between the Beryciformes and the Perciformes. This group is also known as the Zeomorphi or Zeoidea. There is no orbitosphenoid bone; the pelvic fin has a spine and from five to nine soft rays; the simple posttemporal is rigidly united to the skull; and there is a more or less distinct anterior, spinous anal fin of one t four spines, as well as a spinous dorsal fin. Th zeiform fishes, which are known from the Paleocen



John Dory, *Zenopsis ocellata*. (After G. B. Goode, Great International Fisheries Exhibition, London, 1883, U S Natl Museum Bull 27, 1884)

on, are grouped into 3 families, perhaps 12 genera, and fewer than 50 species. All are marine, living in shore waters and chiefly at moderate depths off tropical and temperate coasts. Most are of small size and they are of minor economic importance. See ACTINOPTERYGII. [R.M B]

Zenith

The point directly overhead in the sky. The astronomical zenith, which is that usually meant, is the upper intersection of a plumb line with the celestial sphere. The zenith distance of a celestial object is its angular distance from the astronomical zenith, and is identical with the complement of its altitude. The geocentric zenith is the upper intersection with the celestial sphere of an imagined line through the center of Earth and the observer. The point diametrically opposite the zenith is called the nadir. See ASTRONOMICAL COORDINATE SYSTEMS. [G M.C.]

Zeolite

Any mineral belonging to the closely related group of hydrous aluminum silicates known as the zeolite family. There are nearly 40 zeolites, many of which are rare and only a few—notably natrolite, analcime, stilbite, heulandite, and chabazite—are common (see table). They all have a similar occurrence and are usually found as secondary minerals in cracks and cavities in basic igneous rocks, chiefly basalt. Zeolites are characteristically found in good crystals associated with each other and with calcite, datolite, apophyllite, and prehnite. The crystals are usually colorless. They have a low specific gravity, 2.0-2.4, and hardness is $3\frac{1}{2}$-$5\frac{1}{2}$ on Mohs scale. All zeolites are easily decomposed by hydrochloric

acid When placed in the flame they appear to boil as they fuse to a blebby glass

the feldspars they are tectosilicates built on a three-dimensional framework of linked SiO_4 tetrahedra in which aluminum takes the place of some of the silicon The excess negative charge resulting from the substitution of Al for Si is compensated for by the introduction in the structure of Ca, Na, and K *See* SILICATE MINERALS.

The crystallography of the zeolites differs greatly from one to another, since they crystallize in the isometric, orthorhombic, monoclinic, and hexagonal systems Nevertheless, all have similar open structures and contain channelways in which water is readily housed This water differs from that contained in most minerals, for as the zeolite is heated, the water is given off continuously, leaving the structure intact. After nearly complete dehydration, the channels may again be filled with water or other materials such as ammonia, iodine, or alcohol This process is selective depending on the relative sizes of the molecules and the channelways. In a given zeolite certain large molecules will be excluded whereas smaller ones will be admitted; hence zeolites are used as molecular sieves Crystalline substances resembling natural zeolites are manufactured for this purpose *See* MOLECULAR SIEVE

Zeolites have another interesting property based on the structure, that of base exchange or cation exchange Through this agency zeolites or synthetic substances with zeolitic structure are used

Some minerals in the zeolite family

Natrolite group	
Natrolite	$Na_2[Al_2Si_3O_{10}]\,2H_2O$
Mesolite	$Na_2Ca_2[Al_2Si_3O_{10}]_3\,8H_2O$
Thomsonite	$NaCa_2[Al_2(Al,Si)Si_2O_{10}]_2\,5H_2O$
Scolecite	$Ca[Al_2Si_3O_{10}]\,3H_2O$
Laumontite-gismondite group	
Gismondite	$Ca[Al_2Si_2O_8]\,4H_2O$
Laumontite	$Ca[AlSi_2O_6]_2\,4H_2O$
Mordenite	$(Ca,K_2,Na_2)[AlSi_5O_{12}]_2\,7H_2O$
Heulandite-stilbite group	
Heulandite	$Ca[Al_2Si_7O_{18}]\,6H_2O$
Stilbite	$Ca[Al_2Si_7O_{18}]\,7H_2O$
Epistilbite	$Ca[Al_2Si_6O_{16}]\,5H_2O$
Brewsterite	$(Sr,Ba,Ca)[Al_2Si_6O_{16}]\,5H_2O$
Phillipsite group	
Phillipsite	$KCa[Al_3Si_5O_{16}]\,6H_2O$
Harmotome	$Ba[Al_2Si_6O_{16}]\,6H_2O$
Chabazite group	
Gmelinite	$(Na_2,Ca)[Al_2Si_4O_{12}]\,6H_2O$
Chabazite	$(Ca,Na_2)[Al_2Si_4O_{12}]\,6H_2O$
Analcime*	$Na[AlSi_2O_6]\,H_2O$

* Chemically analcime is closely related to leucite and is sometimes considered a member of the feldspathoid group

as water softeners This is accomplished by passing hard water through a container filled with grains of a sodium zeolite The calcium ions of the water exchange places with the sodium ions of the zeolite, making the water soft. When the zeolite has had all the sodium removed, it can be regenerated by passing a sodium chloride brine through the container and forcing the exchange in the opposite direction *See* WATER SOFTENING. [C.S.HU.]

Zero

In mathematics, the concept zero is used in two ways: as a *number* and as a value of a variable. The positional system of number notation, developed first by the Babylonians with the basis 60, and later by the Hindus and the Chinese with the base 10,

the Arabs as the arabic number system

The zero as a number, however, is a new concept, introduced by the Hindus and Chinese about the same time (sixth century). Brahmagupta (born A.D. 598) remarked that the number 0 has special properties: $a \pm 0 = a$, and $a \cdot 0 = 0$, where a may be any number (integer). Using zero as a denominator, he stated $a/0 \pm b = a/0$, where thus $a/0$ appears as a number of a new kind, which cannot be changed by addition or subtraction (In the modern way of thinking, division by zero is therefore not a "permissible" operation) In the acceptance of zero as a number, the Orient is centuries ahead of the Occident Not until Leonardo Fibonacci (*Liber abaci*, 1228) was the zero treated. It is of interest that the negative numbers appear earlier in the history of mathematics than does the number zero; they were already known to the Chinese mathematician Liu Hui in the third century A.D.

In a modern way, zero can be called the identity element of the infinite Abelian additive group of integers. If in an integral domain (*a fortiori* in a field) a product is equal to zero then at least one factor of the product is zero

In the second concept zero is the value of a variable for which a function is equal to zero. For example, "A polynomial of degree n has n zeros," or "The Riemann zeta function $\zeta(s)$ has all its complex zeros in the strip $0 <$ *real part* $s < 1$" *See* NUMBER THEORY. [H R]

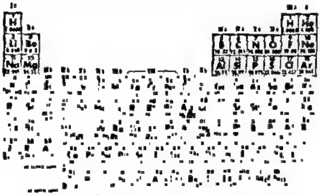
Zinc

Chemical element number 30, zinc, Zn, is a malleable, ductile, gray metal with atomic weight 65 38 Because of chemical similarities among zinc, cadmium, and mercury, these three metals are classed together in group IIb of the periodic table of elements Zinc became known in Europe during the Middle Ages, although it had been known much earlier in Asia, Paracelsus, in the sixteenth century, is said to have been the first European to recognize zinc as a distinct metallic element and to call it "zinckum."

Thirteen isotopes of zinc are known, of which five are stable, having atomic masses of 64, 66, 67,

68, and 70. About half of ordinary zinc occurs as the isotope of atomic mass 64.

Zinc and zinc compounds have long been considered moderately poisonous, but in general they are not. The hydrolysis of zinc compounds produces some irritating acidity, sometimes mistaken for toxicity. Furthermore, traces of cadmium, arsenic, lead, or antimony, sometimes found in zinc, are poisonous enough to make impure zinc something of a hazard. Zinc containers may be used for storing drinking water but must not be used for food.



Uses of the metal. The most important use of *zinc is as a protective coating on other metals.* Coating iron or steel with zinc is called galvanizing, and it may be done by immersing the article in melted zinc (hot dip process), electrolytic deposition onto the article in a plating bath (electrogalvanizing), exposing the article to powdered zinc near its melting point (sherardizing), or spraying the article with melted zinc (metallizing) The mere physical presence of the zinc coat prevents corrosion of iron, and even if breaks in the coat expose portions of the iron, the greater chemical activity of the zinc causes it to be consumed in preference to the iron. Other important uses of zinc are in brass and zinc die-casting alloys, in zinc sheet and strip, in electrical dry cells, in making certain zinc compounds, and as a reducing agent in chemical preparations. Zinc is also an essential element in the growth of many kinds of organisms, both plant and animal The compound insulin is a zinc-containing protein. *See* ELECTROPLATING OF METALS; METAL COATINGS; ZINC ALLOYS.

Occurrence. Zinc is one of the less common elements, making up 0.004% of the earth's crust It is twenty-fifth in order of abundance among the elements. About one-sixth of the zinc ore mined in the world comes from the United States, and nearly as much comes from Canada. The chief ore is zinc blende, or sphalerite, ZnS. Other ores include calamine, $Zn_2SiO_4 \cdot H_2O$; smithsonite, $ZnCO_3$; willemite, Zn_2SiO_4; zincite, ZnO; and franklinite, $(Zn,Fe,Mn)O \cdot (Fe,Mn)_2O_3$. The last two ores are found almost exclusively in the United States. Much zinc is also separated as a by-product of processing iron ores. For a discussion of the commercial production of zinc, *see* ZINC METALLURGY.

Zinc metal. Pure, freshly polished zinc is bluish white and lustrous. Moist air brings about a superficial tarnishing to give the metal its usual grayish color. Pure zinc is malleable and ductile enough to be rolled or drawn, but small amounts of other metals present as contaminants may render it brittle. Zinc melts at 419°C and boils at 906°C. Its density is 7.13 times that of water, so that 1 ft^3 of zinc weighs 445 lb. As a conductor of heat and electricity, zinc ranks high. However, each of its conductance values is only about one-fourth that of silver, the best of the metals in these respects.

Chemical properties. Zinc is a fairly active metal chemically. It can be ignited with some difficulty to give a blue-green flame in air. Zinc ranks above hydrogen in the electrochemical series, so that metallic zinc in an acidic solution will react to liberate hydrogen gas as the zinc passes into solution to form dipositively charged zinc ions, Zn^{++}. This reaction is slow with very pure zinc, but the presence of small amounts of impurities, addition of a trace of copper sulfate, or contact between the zinc surface and such metals as nickel or platinum facilitates formation of gaseous hydrogen and speeds the reaction The combination of zinc and dilute acid is often used to generate small quantities of hydrogen in the laboratory. Zinc also dissolves in strongly alkaline solutions, such as sodium hydroxide, to liberate hydrogen and form dinegatively charged tetrahydroxozincate ions, $Zn(OH)_4^{--}$, sometimes written as ZnO_2^{--} in the formulas of the zincate compounds The common soluble zinc compounds undergo to some extent the process of hydrolysis, which makes their solutions slightly acidic. The ion Zn^{++} is colorless, so that the relatively few zinc compounds that are not colorless in large crystals, or white as powders, receive their color through the influence of the other constituents.

Zinc compounds. Zinc is always divalent in its *compounds, except for some of those with other* metals, which are classed as zinc alloys. The table lists some of the more important zinc compounds.

Compounds of zinc

Name and formula	Uses	Properties, remarks
Zinc oxide, ZnO	Reinforcer and vulcanization activator in rubber tires, pigment, ointment base, in *cements and ceramics*	White, nearly insoluble powder, slightly alkaline in water, turns yellow at 500°C, returns to white *on cooling*, nontoxic and nonirritating; amphoteric, dissolving in either acid or base
Zinc chloride, $ZnCl_2$; Zinc chloride 4-hydrate, $ZnCl_2 \cdot 4H_2O$	Soldering and welding flux, in dye printing, dye manufacture, and wood treating	Colorless, soluble compound, solutions acidic, when made alkaline, forms a series of zinc oxychlorides; exists in several hydrated forms *with less than* $4H_2O$ per molecule

Compounds of zinc (Cont)

Name and formula	Uses	Properties, remarks
Zinc sulfate, $ZnSO_4$	Agriculturally, in arsenical sprays, in rayon manufacture, dyeing, printing, in electrogalvanizing; ingredient in making lithopone pigment	White, soluble compound, commercial form is hydrated
Zinc sulfide, ZnS	Component of lithopone pigment, in luminescent form as a fluorescer or phosphor in television screens, oscilloscopes, x-ray apparatus	White, insoluble; only common white sulfide of a metal, will not blacken in sulfide vapors
Zinc chromate, $ZnCrO_4$	In rust-inhibiting coatings	Usually a combined zinc hydroxide-chromate such as $ZnCrO_4 \cdot 4Zn(OH)_2$, sometimes contains potassium also, varying shades of yellow
Zinc orthosilicate, Zn_2SiO_4	In fluorescers, as a refractory, in water softeners	Used as both anhydrous form (*willemite*) and 1-hydrate (calamine)
Sodium zincate, Na_2ZnO_2	Water softener, paper- and cloth-treating agent, flocculating agent in water purification	Formed in strongly alkaline solution
Zinc fluoride, ZnF_2	Catalyst, wood preservative, termite repellent	Rather low solubility in water
Zinc dithionite, ZnS_2O_4	Bleaching agent for wood pulp, rags, and clay	Also called zinc hydrosulfite
Zinc peroxide, ZnO_2	Prophylactic against wound infections	Commercially about half ZnO_2, half ZnO
Zinc orthophosphate 3-hydrate, $ZnHPO_4 \cdot 3H_2O$	Dental cement	Anhydrous form also known, as are many other zinc phosphates

Zinc also forms many *coordination compounds* The *zincates* are actually coordination compounds, or complexes, *in which hydroxide* ions, OH^-, are bound to the zinc ions. Ammonia, NH_3, forms complexes with zinc, such as the typical tetrammine zinc ion, $[Zn(NH_3)_4]^{++}$ Zinc cyanide, usually given the simple formula $Zn(CN)_2$, is a coordination compound in which many alternating zinc and cyanide ions are three-dimensionally bound together in a very large molecule. In most coordination compounds of zinc, the fundamental structural unit is a central zinc ion surrounded by four coordinated groups arranged spatially at the corners of a regular tetrahedron

Analytical methods. White zinc sulfide is precipitated by *ammonium sulfide or hydrogen sulfide* from neutral or alkaline zinc salt solutions The precipitate is distinguished from those of nickel and cobalt by its lack of color and by its solubility in acid Sodium, potassium, or ammonium hydroxide precipitates white zinc hydroxide. The solubility of the hydroxide in excess sodium hydroxide distinguishes it from those of iron and manganese, and the solubility in excess ammonium *hydroxide from those of aluminum* and chromium as well. Qualitative identification may also be made by ignition of a zinc compound in the presence of cobalt nitrate, which yields a greenish solid.

Quantitatively, zinc may be determined by precipitating the sulfide, igniting it to form the oxide, and weighing Zinc *ammonium orthophosphate*, $ZnNH_4PO_4$, may be precipitated, ignited, and *weighed as the pyrophosphate*, $Zn_2P_2O_7$. Quantitative electrolysis of zinc salts in acetate buffer solution is possible, with the zinc being plated out and weighed as metallic zinc. A volumetric method for zinc salts in solution involves titrating with standard potassium ferrocyanide solution *See* CADMIUM, MERCURY (ELEMENT). [W E C.]

Bibliography: J C. Bailar, Jr. (ed.), *The Chemistry of the Coordination Compounds*, 1956, H. M. Cyr, M C. Sneed, and R. C. Brasted (eds.), *Comprehensive Inorganic Chemistry*, vol 4, 1955; N. A. Lange and G M. Forker, *Handbook of Chemistry*, 9th ed., 1956, C H Mathewson (ed), *Zinc*, ACS Monograph 142, 1959, R. J Meyer (ed.), *Gmelin's Handbuch der Anorganischen Chemie*, 10(32), 1956; N V Sidgwick, *The Chemical Elements and their Compounds*, vol 1, 1950

Zinc alloys

Combinations of zinc with one or more other metals *Zinc crystallizes* in the close-packed hexagonal system, and the solid solubility of most of the common cubic metals is limited to a very small percentage Attempts to form alloys with most metals beyond these limits of solid solubility give rise to the formation of brittle constituents which *render* the alloy useless The outstanding exceptions are aluminum and copper

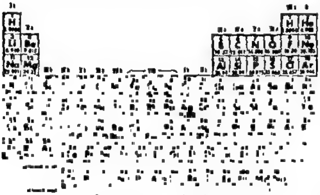
The most important group of zinc alloys are those containing 4% aluminum, together with a few hundredths of 1% magnesium, and in some instances up to 1% *copper*. These alloys are used in die casting and account for about 40% of the total *zinc consumption* in the United States. These alloys must be made with zinc of 99.99% purity, *meeting* the ASTM special high-grade zinc specification

These alloys can be cast [illegible]

can be cast to very close dimensional tolerances, in very complex shapes, and where necessary, with

68, and 70. About half of ordinary zinc occurs as the isotope of atomic mass 64.

Zinc and zinc compounds have long been considered moderately poisonous, but in general they are not. The hydrolysis of zinc compounds produces some irritating acidity, sometimes mistaken for toxicity. Furthermore, traces of cadmium, arsenic, lead, or antimony, sometimes found in zinc, are poisonous enough to make impure zinc something of a hazard. Zinc containers may be used for storing drinking water but must not be used for food.



Uses of the metal. The most important use of zinc is as a protective coating on other metals. Coating iron or steel with zinc is called galvanizing, and it may be done by immersing the article in melted zinc (hot dip process), electrolytic deposition onto the article in a plating bath (electrogalvanizing), exposing the article to powdered zinc near its melting point (sherardizing), or spraying the article with melted zinc (metallizing). The mere physical presence of the zinc coat prevents corrosion of iron, and even if breaks in the coat expose portions of the iron, the greater chemical activity of the zinc causes it to be consumed in preference to the iron. Other important uses of zinc are in brass and zinc die-casting alloys, in zinc sheet and strip, in electrical dry cells, in making certain zinc compounds, and as a reducing agent in chemical preparations. Zinc is also an essential element in the growth of many kinds of organisms, both plant and animal. The compound insulin is a zinc-containing protein. See ELECTROPLATING OF METALS; METAL COATINGS; ZINC ALLOYS.

Occurrence. Zinc is one of the less common elements, making up 0.004% of the earth's crust. It is twenty-fifth in order of abundance among the elements. About one-sixth of the zinc ore mined in the world comes from the United States, and nearly as much comes from Canada. The chief ore is zinc blende, or sphalerite, ZnS. Other ores include calamine, $Zn_2SiO_4 \cdot H_2O$; smithsonite, $ZnCO_3$; willemite, Zn_2SiO_4; zincite, ZnO; and franklinite, $(Zn,Fe,Mn)O \cdot (Fe,Mn)_2O_3$. The last two ores are found almost exclusively in the United States. Much zinc is also separated as a by-product of processing iron ores. For a discussion of the commercial pro-

... bluish ... super-
ficial tarnishing to give the metal its usual grayish color. Pure zinc is malleable and ductile enough to be rolled or drawn, but small amounts of other metals present as contaminants may render it brittle. Zinc melts at 419°C and boils at 906°C. Its density is 7.13 times that of water, so that 1 ft³ of zinc weighs 445 lb. As a conductor of heat and electricity, zinc ranks high. However, each of its conductance values is only about one-fourth that of silver, the best of the metals in these respects.

Chemical properties. Zinc is a fairly active metal chemically. It can be ignited with some difficulty to give a blue-green flame in air. Zinc ranks above hydrogen in the electrochemical series, so that metallic zinc in an acidic solution will react to liberate hydrogen gas as the zinc passes into solution to form dipositively charged zinc ions, Zn^{++}. This reaction is slow with very pure zinc, but the presence of small amounts of impurities, addition of a trace of copper sulfate, or contact between the zinc surface and such metals as nickel or platinum facilitates formation of gaseous hydrogen and speeds the reaction. The combination of zinc and dilute acid is often used to generate small quantities of hydrogen in the laboratory. Zinc also dissolves in strongly alkaline solutions, such as sodium hydroxide, to liberate hydrogen and form dinegatively charged tetrahydroxozincate ions, $Zn(OH)_4^{--}$, sometimes written as ZnO_2^{--} in the formulas of the zincate compounds. The common soluble zinc compounds undergo to some extent the process of hydrolysis, which makes their solutions slightly acidic. The ion Zn^{++} is colorless, so that the relatively few zinc compounds that are not colorless in large crystals, or white as powders, receive their color through the influence of the other constituents.

Zinc compounds. Zinc is always divalent in its compounds, except for some of those with other metals, which are classed as zinc alloys. The table lists some of the more important zinc compounds.

Compounds of zinc

Name and formula	Uses	Properties, remarks
Zinc oxide, ZnO	Reinforcer and vulcanization activator in rubber tires, pigment, ointment base, in cements and ceramics	White, nearly insoluble powder; slightly alkaline in water; turns yellow at 500°C, returns to white on cooling, nontoxic and nonirritating, amphoteric, dissolving in either acid or base
Zinc chloride, $ZnCl_2$; Zinc chloride 4-hydrate, $ZnCl_2 \cdot 4H_2O$	Soldering and welding flux, in dye printing, dye manufacture, and wood treating	Colorless, soluble compound, solutions acidic; when made alkaline, forms a series of zinc oxychlorides, exists in several hydrated forms with less than $4H_2O$ per molecule

Compounds of zinc (Cont.)

Name and formula	Uses	Properties, remarks
Zinc sulfate, $ZnSO_4$	Agriculturally, in arsenical sprays, in rayon manufacture, dyeing, printing, in electrogalvanizing; ingredient in making lithopone pigment	White, soluble compound, commercial form is hydrated
Zinc sulfide, ZnS	Component of lithopone pigment; in luminescent form as a fluorescer or phosphor in television screens, oscilloscopes, x-ray apparatus	White, insoluble, only common white sulfide of a metal, will not blacken in sulfide vapors
[illegible] chromate, [illegible]	In rust-inhibiting coatings	Usually a combined zinc hydroxide-chromate such as $ZnCrO_4 \cdot 4Zn(OH)_2$, sometimes contains potassium also, varying shades of yellow
[illegible] orthosilicate, [illegible]	In fluorescers, as a refractory; in water softeners	Used as both anhydrous form (willemite) and 1-hydrate (calamine)
[illegible] zincate, [illegible]	Water softener, paper- and cloth-treating agent, flocculating agent in water purification	Formed in strongly alkaline solution
Zinc fluoride, ZnF_2	Catalyst, wood preservative, termite repellent	Rather low solubility in water
Zinc [illegible], [illegible]	Bleaching agent for wood pulp, rags, and clay	Also called zinc hydrosulfite
Zinc peroxide, [illegible]	Prophylactic against wound infections	Commercially about half ZnO_2, half ZnO
Zinc phosphate [illegible] hydrate, [illegible]	Dental cement	Anhydrous form also known, as are many other zinc phosphates

Zinc also forms many coordination compounds. The zincates are actually coordination compounds, complexes in which hydroxide ions, OH^-, are bound to the zinc ions. Ammonia, NH_3, forms complexes with zinc, such as the typical tetrammine zinc ion, $[Zn(NH_3)_4]^{2+}$. Zinc cyanide, usually given the simple formula $Zn(CN)_2$, is a coordination compound in which many alternating zinc and cyanide ions are three-dimensionally bound together in a very large molecule. In most coordination compounds of zinc, the fundamental structural unit is a central zinc ion surrounded by four coordinated groups arranged spatially [illegible]

Analytical methods. White zinc sulfide is precipitated by ammonium sulfide or hydrogen sulfide from neutral or alkaline zinc salt solutions. The precipitate is distinguished from those of nickel and cobalt by its lack of color and by its solubility in acid. Sodium, potassium, or ammonium hydroxide precipitates white zinc hydroxide. The solubility of the hydroxide in excess sodium hydroxide distinguishes it from those of iron and manganese, and the solubility in excess ammonium hydroxide from those of aluminum and chromium as well. Qualitative identification may also be made by ignition of a zinc compound in the presence of cobalt nitrate, which yields a greenish solid.

Quantitatively, zinc may be determined by precipitating the sulfide, igniting it to form the oxide, and weighing. Zinc ammonium orthophosphate, $ZnNH_4PO_4$, may be precipitated, ignited, and weighed as the pyrophosphate, $Zn_2P_2O_7$. Quantitative electrolysis of zinc salts in acetate buffer solution is possible, with the zinc being plated out and weighed as metallic zinc. A volumetric method for zinc salts in solution involves titrating with standard potassium ferrocyanide solution. See CADMIUM; MERCURY (ELEMENT). [W.E.C.]

Bibliography: J. C. Bailar, Jr. (ed.), *The Chemistry of the Coordination Compounds*, 1956; H. M. Cyr, M. C. Sneed, and R. C. Brasted (eds.), *Comprehensive Inorganic Chemistry*, vol. 4, 1955; N. A. Lange and G. M. Forker, *Handbook of Chemistry*, 9th ed., 1956; C. H. Mathewson (ed.), *Zinc*, ACS Monograph 142, 1959; R. J. Meyer (ed.), *Gmelin's Handbuch der Anorganischen Chemie*, 10(32), 1956; N. V. Sidgwick, *The Chemical Elements and their Compounds*, vol. 1, 1950.

Zinc alloys

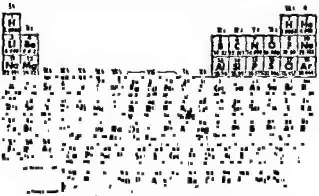
Combinations of zinc with one or more other metals. Zinc crystallizes in the close-packed hexagonal system, and the solid solubility of most of the common cubic metals is limited to a very small percentage. Attempts to form alloys with most metals beyond these limits of solid solubility give rise to the formation of brittle constituents which render the alloy useless. The outstanding exceptions are aluminum and copper.

The most important group of zinc alloys are those containing 4% aluminum together with a few hundredths of 1% magnesium and in some instances up to 1% copper. These alloys are used in die casting and account for about [illegible] of the total zinc consumption in the United States. [illegible] alloys must be made with zinc of 99.99% [illegible] meeting the ASTM special high-grade [illegible]cation.

These alloys can be cast in the [illegible] die-casting machine, using [illegible] simple steels which [illegible]

68, and 70. About half of ordinary zinc occurs as the isotope of atomic mass 64.

Zinc and zinc compounds have long been considered moderately poisonous, but in general they are not. The hydrolysis of zinc compounds produces some irritating acidity, sometimes mistaken for toxicity. Furthermore, traces of cadmium, arsenic, lead, or antimony, sometimes found in zinc, are poisonous enough to make impure zinc something of a hazard. Zinc containers may be used for storing drinking water but must not be used for food.



Uses of the metal. The most important use of zinc is as a protective coating on other metals. Coating iron or steel with zinc is called galvanizing, and it may be done by immersing the article in melted zinc (hot dip process), electrolytic deposition onto the article in a plating bath (electrogalvanizing), exposing the article to powdered zinc near its melting point (sherardizing), or spraying the article with melted zinc (metallizing). The mere physical presence of the zinc coat prevents corrosion of iron, and even if breaks in the coat expose portions of the iron, the greater chemical activity of the zinc causes it to be consumed in preference to the iron. Other important uses of zinc are in brass and zinc die-casting alloys, in zinc sheet and strip, in electrical dry cells, in making certain zinc compounds, and as a reducing agent in chemical preparations. Zinc is also an essential element in the growth of many kinds of organisms, both plant and animal. The compound insulin is a zinc-containing protein. See ELECTROPLATING OF METALS; METAL COATINGS; ZINC ALLOYS.

Occurrence. Zinc is one of the less common elements, making up 0.004% of the earth's crust. It is twenty-fifth in order of abundance among the elements. About one-sixth of the zinc ore mined in the world comes from the United States, and nearly as much comes from Canada. The chief ore is zinc ... sphalerite, ZnS. Other ores include cala-
$ZnCO_3$; willem-
anklinite, (Zn,Fe,
o ores are found
almost exclusively in the United States. Much zinc is also separated as a by-product of processing iron ores. For a discussion of the commercial production of zinc, see ZINC METALLURGY.

Zinc metal. Pure, freshly polished zinc is bluish white and lustrous. Moist air brings about a superficial tarnishing to give the metal its usual grayish color. Pure zinc is malleable and ductile enough to be rolled or drawn, but small amounts of other metals present as contaminants may render it brittle. Zinc melts at 419°C and boils at 906°C. Its density is 7.13 times that of water, so that 1 ft^3 of zinc weighs 445 lb. As a conductor of heat and electricity, zinc ranks high. However, each of its conductance values is only about one-fourth that of silver, the best of the metals in these respects.

Chemical properties. Zinc is a fairly active metal chemically. It can be ignited with some difficulty to give a blue-green flame in air. Zinc ranks above hydrogen in the electrochemical series, so that metallic zinc in an acidic solution will react to liberate hydrogen gas as the zinc passes into solution to form dipositively charged zinc ions, Zn^{++}. This reaction is slow with very pure zinc, but the presence of small amounts of impurities, addition of a trace of copper sulfate, or contact between the zinc surface and such metals as nickel or platinum facilitates formation of gaseous hydrogen and speeds the reaction. The combination of zinc and dilute acid is often used to generate small quantities of hydrogen in the laboratory. Zinc also dissolves in strongly alkaline solutions, such as sodium hydroxide, to liberate hydrogen and form dinegatively charged tetrahydroxozincate ions, $Zn(OH)_4^{--}$, sometimes written as ZnO_2^{--} in the formulas of the zincate compounds. The common soluble zinc compounds undergo to some extent the process of hydrolysis, which makes their solutions slightly acidic. The ion Zn^{++} is colorless, so that the relatively few zinc compounds that are not colorless in large crystals, or white as powders, receive their color through the influence of the other constituents.

Zinc compounds. Zinc is always divalent in its compounds, except for some of those with other metals, which are classed as zinc alloys. The table lists some of the more important zinc compounds.

Compounds of zinc

Name and formula	Uses	Properties, remarks
Zinc oxide, ZnO	Reinforcer and vulcanization activator in rubber tires, pigment, ointment base, in cements and ceramics	White, nearly insoluble powder; slightly alkaline in water, turns yellow at 500°C, returns to white on cooling; nontoxic and nonirritating, amphoteric, dissolving in either acid or base
Zinc chloride, $ZnCl_2$; Zinc chloride 4-hydrate, $ZnCl_2 \cdot 4H_2O$	Soldering and welding flux; in dye printing, dye manufacture, and wood treating	Colorless, soluble compound, solutions acidic, when made alkaline, forms a series of zinc oxychlorides, exists in several hydrated forms with less than $4H_2O$ per molecule

ously from the bottom. The zinc vapor-gas mixture is condensed to metal or alternatively may be burned to pigment grade oxide A single furnace will produce 50 tons or more of zinc per day

Blast furnace process. The most recent process for extracting zinc is the blast furnace process, developed by the Imperial Smelting Corporation. The heat of reaction is generated by burning carbon in the reaction chamber, which is a short blast furnace The gas has as a result a rather high carbon dioxide content, but is nevertheless successfully condensed by shock cooling with molten lead.

Electrolytic process A large amount of zinc is produced by electrolysis from a sulfate solution The roasted concentrate is leached with the spent electrolyte from the cells, thus regenerating the zinc sulfate which, after rigorous purification, is returned to the electrolytic cells Lead alloy anodes are used; the cathodes [illegible]

[illegible] agents to the electrolyte, coupled with a very high degree of purification of the electrolyte, permit the direct production of high-purity zinc

Refining. The zinc produced by all of the carbon reduction processes requires a refining step to produce zinc of the highest grade A process widely employed commercially is based on fractional distillation in reflux refining columns. A typical unit consists of two lead columns followed by a single cadmium column, the lead column serving [illegible]

Commercial grades of zinc. The American Society for Testing Materials recognizes six grades of zinc, varying in purity The purest grade, which is known as special high-grade, has a purity in excess of 99.99% zinc. This grade is produced commercially either by the refining process described above or by the electrolytic process. Approximately 35% of the zinc consumed in the United States is of this grade *See* PYROMETALLURGY, NONFERROUS [W.M.F.]

Zincate

A negative ion, usually given the formula ZnO_2^{2-}, is derived from zinc hydroxide Zinc hydroxide is an amphoteric substance and thus can react with a strong base such as sodium hydroxide to form sodium zincate,

$$Zn(OH)_2 + 2NaOH \rightarrow Na_2ZnO_2 + 2H_2O$$

and also react in the more usual manner with acids to form zinc salts

Zincate solutions are strongly basic, and are used to plate zinc on aluminum prior to copper plating *See* AMPHOTERISM; ZINC. [E.E.WR.]

Zincite

A mineral with composition ZnO (zinc oxide) Although it crystallizes in the hexagonal system in pyramidal crystals, these are rare and it is usually massive Cleavage is prismatic Its hardness is 4 and its specific gravity 5.6. The mineral has a subadamantine luster and deep red to orange-yellow color. Manganese is present in varying amounts and probably colors the mineral, for pure ZnO is [illegible]

as a valuable ore of zinc. *See* ZINC. [C.S.HU.]

Zipper

A generic name for slide fasteners that provide sturdy and continuous closure for adjacent pieces of textile, leather, rubber, or plastic materials. The first real slide fastener was patented in 1893 by W. L. Judson of Chicago. The most common type, in which the two sets of interlocking teeth are of the same design on both sides of the material, was patented in Germany and elsewhere by Catharina Kuhn-Moos in 1912. Modern zippers, of which there are more than 130 styles, are made of brass, steel, aluminum, nickel-silver, and even nylon. In the 1950s, a tongue-and-groove all-plastic zipper, differing in principle but identical in function, came into use on articles such as brief cases and phonograph record covers *See* TEXTILE [C.CO.]

Zircon

A mineral with the idealized composition $ZrSiO_4$, one of the chief sources of the element zirconium Trace amounts of uranium and of thorium are often present and the mineral may then be partly or entirely metamict The name cyrtolite is applied to an altered type of zircon Structurally, zircon is a nesosilicate, with isolated SiO_4 groups. It is isostructural with the thorium silicate thorite and the yttrium phosphate xenotime. *See* SILICATE MINERALS, RADIOACTIVE MINERALS; ZIRCONIUM.

Zircon is tetragonal in crystallization. It often occurs as well-formed crystals, which commonly are square prisms terminated by a low pyramid. The color is variable, usually brown to reddish brown, but also colorless, pale yellowish, green, or blue The transparent colorless or tinted varieties are popular gem stones. Hardness is 7½ on Mohs scale; specific gravity is 4.7, decreasing in metamict types

Because of its chemical and physical stability, zircon resists weathering and accumulates in residual deposits and in beach and river sands, from which it has been obtained commercially in Florida and in India, Brazil, and other countries. It is widespread in small amounts as an accessory mineral in granitic and syenitic igneous rocks, and occurs more abundantly in pegmatite deposits associated with such rocks. Zircon also is a minor constituent in many types of gneissic and schistose metamorphic rocks [C.FR.]

Zirconium

A metallic element, symbol Zr, atomic number 40, atomic weight 91.22 (natural isotopes 90, 91, 92, 94, 96)

Occurrence. Zirconium is one of the more abundant elements, and it is distributed widely in the

very thin sections. They may be finished by electroplating or other methods where desired.

The naturally occurring impurities in zinc which may be looked upon as alloying elements are lead and cadmium; for the other major use of zinc, the zinc coating of steel (galvanizing), lead and cadmium contents of the zinc employed are regulated by selection or blending of the grades normally produced by smelting. The purpose of such control of composition is primarily to obtain a desired appearance (spangle) on the galvanized product. *See* METAL COATINGS

About 5% of the zinc consumed is in the form of rolled strips or sheets Rolled zinc is used in the manufacture of dry batteries and here, as in the case of galvanizing, the composition with respect to the natural impurities, lead, cadmium, and in this case iron, is controlled by selection of normally produced grades.

Rolled zinc is also used for roofing, flashing, leaders, and gutters; alloyage of copper up to 1% and very small amounts of other metals which give rise to a dispersion-hardening constituent, for example, titanium, may be used.

Zinc is an important alloying element in the copper-base alloys. These alloys, the brasses, may contain up to 40% zinc Zinc is also used as an alloying element, in smaller amounts, in certain aluminum alloys and in nickel silver. *See* CASTING; COPPER ALLOYS; ZINC; ZINC METALLURGY.

[W.M.P.]

Zinc metallurgy

The removal from ores, refining, and preparation of zinc. By far the largest group of ores from which zinc is extracted is the sulfides. The concentrates shipped to the smelters typically contain 50–60% zinc, substantially all of which is present as zinc sulfide. A minor source of zinc is oxidized ores in which the zinc is present predominantly as zinc oxide, zinc silicate, or zinc carbonate The usual impurities are iron (pyrite), lead, and cadmium

Roasting. Sulfide ores must be roasted to convert the zinc to zinc oxide prior to smelting The roasting step may be carried out in various types of rabbled hearth furnaces, but in modern practice the roasting is usually done in suspension (flash) roasters in which the finely ground zinc sulfide concentrate is blown into a combustion chamber and burned, much as powdered coal is burned, or the roasting may be carried out in some type of fluidized-bed roaster.

In either case, the sulfur content of the calcine is reduced to a low value and a gas is generated containing a concentration of sulfur dioxide suitable for conversion to sulfuric acid. The production of sulfuric acid is mandatory in many locations to avoid the air pollution problem which would otherwise result.

Fire concentration. Low-grade ores of the oxide type and secondary-zinc-bearing materials may under some conditions be concentrated economically in a rotary kiln operation known as the waelz process. The zinc-bearing material, mixed with sufficient coal to reduce the zinc oxide, is charged into a rotary kiln. In the lower (discharge) end of the kiln, the temperature necessary for reduction and vaporization of the zinc is reached. As the gas stream carrying the zinc vapor travels up the kilns, the zinc vapor is reoxidized by excess air, and the heat generated is partially transferred to the entering charge. The zinc oxide entrained in the gas stream is collected in a bag house and sent to the smelter.

Sintering. For most pyrometallurgical extraction methods, the roasted concentrates or the waelz oxide must be sintered to bring them to a suitable density and particle size. This sintering is usually carried out on a Dwight Lloyd sintering machine, but may be carried out in a rotary kiln.

Smelting methods. All pyrometallurgical processes and electrothermic processes are based on the highly endothermic reaction by which carbon reduces zinc oxide at a temperature of 1000°C or higher, which is above the boiling point of zinc. Sufficient carbon must be present to prevent oxidation of the carbon beyond carbon monoxide, since any large percentage of carbon dioxide will result in reoxidation of the zinc vapor during the condensation step.

All but one of the carbon reduction processes are retort processes in which the heat is generated outside the retort in which the reaction occurs.

Horizontal (Belgian) retort process. The earliest process, which is still widely used, employs clay or silicon carbide-clay retorts, 8–10 in. in inside diameter and 5–6 ft long, set horizontally in a coal- or gas-fired furnace. Several hundred retorts may be set in one furnace, arranged in 4–8 horizontal rows. The loose charge of roasted concentrate mixed with the necessary amount of coal is charged into the retort. When the reaction temperature is reached, a mixture of zinc vapor and carbon monoxide issues from the mouth of the retort and is condensed in a clay vessel which is luted to the retort Zinc is tapped by hand from these condensers three or four times during a cycle, which may be a 24- or a 48-hour cycle

Vertical retort (New Jersey) process In this process a vertical retort of silicon carbide brick, having a horizontal cross section of 1 ft by 6–8 ft and a vertical heated height of as much as 35 ft, is employed The long walls are heated externally by gas A briquetted charge of coal and roasted concentrate is coked and then charged at the top at frequent intervals, and the residue is withdrawn continuously from the bottom The zinc vapor-carbon monoxide gas is conducted to a condenser through an offtake near the top of the retort Such a retort may have a capacity of over 10 tons per day, as compared with 25–30 lb from a single horizontal retort

Electrothermic (St Joe) process In this process a mixture of carefully sized, hard-sintered, roasted concentrate and coke is charged to the top of a large cylindrical furnace in which the heat of reaction is supplied by the electrical resistance of the charge itself. The residue is withdrawn continu-

where T is in °K and p in mm of mercury. Its electrical resistivity (40 micro ohm-cm at 0°C) is about 25 times that of copper and about one-half that of Nichrome. Its cross section for absorption of neutrons is 0.18 barns (1 barn = 10^{-24} cm²) compared to the following values for other structural materials: Fe, 2.4; Ni, 4.5; Cu, 3.5; Al, 0.22; Mg 0.06.

The metal does not react with the common gases at room temperature and it remains bright and shiny in air indefinitely. At elevated temperatures it is very reactive to all but the inert gases, reacting in two ways: (1) to form a separate phase or compound (ZrO_2, ZrN), and (2) to dissolve large amounts of the gas to form solid solutions that still maintain the structure and some of the properties of the metal. At 700°C as much as 30 atom % oxygen, 20 atom % nitrogen, and 50 atom % hydrogen will dissolve in the metal. The metal reacts with the halogens at 200°C to form only the normal tetrahalides. Finely divided zirconium (powder or chips) is pyrophoric and hazardous to handle and store. Spontaneous explosions have been reported for both the wetted sponge and stored scrap (wet or dry).

The resistance of the metal to corrosion by *aqueous solutions* depends somewhat on conditions but *in general is very good*. Its resistance to alkalies is better than that of tantalum, titanium, or stainless steels. Its resistance to acids is very good except for hydrofluoric acid or very concentrated sulfuric or phosphoric acids. In spite of this resistance to acids and alkalies, its corrosion resistance to *pure water* and steam is not good. The alloy Zircalloy-2 (Zr with 1.5% Sn, 0.12% Fe, 0.05% Ni, 0.10% Cr) with good *resistance to water* and steam *has been developed*.

Compounds. Zirconium is a *member of* subgroup IVb of the periodic table and has a *normal valence of 4*. In some compounds, such as the high-melting ZrO_2, the bonds are largely ionic, whereas in others its bonds are *at least partly covalent*, as in the relatively low-melting halides. *The degree of ionic character in the tetrachlorides of the* group IVb elements increases with *atomic number* as evidenced by their *melting points*.

Tetrahalides of group IVb elements

	Melting point, °C	Boiling point, °C
$TiCl_4$	−30	136
$ZrCl_4$	437	331 (sublimes)
$HfCl_4$	432	317 (sublimes)
$ThCl_4$	820	720–750 (sublimes)

The tetrapositive zirconium ion (Zr^{4+}) does not exist in aqueous solution. Because of the high charge and small size of the ion it combines strongly with the negative ions. The simplest ion formed in *solution* is the zirconyl ion (ZrO^{++}). In sulfate solutions the zirconium is found in negatively charged ions, $ZrO(SO_4)_2^{--}$, and in complex chains of ZrO^{++} and SO_4^{--} links.

Zirconium tetrachloride can be prepared by the action of *chlorine on the carbide and similar* compounds, or on the metal. It is very hygroscopic, and *it reacts violently* with water to form zirconyl chloride ($ZrOCl_2$), which dissolves readily and hydrolyzes to give very acid solutions. The di- and trichloride have been made by reduction of the tetrachloride with aluminum at elevated temperatures. The bromides and iodides are very similar to the chlorides.

The sulfates of zirconium are numerous and complex. They range from the *simple sulfates*, $Zr(SO_4)_2$ and $Zr(SO_4)_2 \cdot 4H_2O$; through the sulfatozirconylates, $H_4ZrO(SO_4)_3$ and $H_2ZrO_2SO_4$; to *the chain compounds, such as* $HO(OZrSO_4)_3ZrOOH \cdot 14H_2O$.

Zirconium dioxide or zirconia has three crystal forms: the monoclinic (stable below 1000°C); the tetragonal (stable from 1000 to 1900°C); and the cubic (stable above 1900°C), which melts at 2680°C. When subjected to prolonged heating, zirconia tends to crack because of conversion *from* one crystal form to another. By *incorporating about* 5% CaO or MgO *such cracking is avoided* because the zirconia remains in the cubic form. This product *is called stabilized zirconia*.

Zirconium can absorb hydrogen to form several phases. The *commercial zirconium hydride has* [illegible] approximately [illegible]

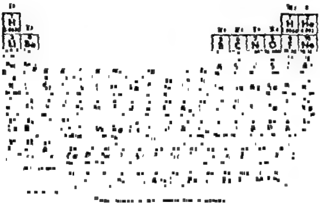
On heating, it loses its hydrogen and leaves powdered zirconium. It is used in powder metallurgy and as a getter in vacuum tubes.

In certain compounds the valence of zirconium is less than four, for example, ZrN, ZrB. In these *compounds and also in ZrC* the structural arrangement of the zirconium atoms is essentially that of *the metal*, whereas the nonmetal atoms are located in sites between the zirconium atoms. The stoichiometry (that is, the ratio of number of metal atoms to nonmetal atoms) of these compounds is determined by the number of vacant sites in the zirconium structure rather than by the normal valence rules. These interstitial compounds are marked by *high melting* points and great hardness (8–9 *on* the Mohs scale). Their electrical conductivities are fairly good, and they *increase with temperature*, as do those of metals. The melting points of some typical interstitial compounds are *given* in the list.

Compound	*Melting point, °C*	*Compound*	*Melting point, °C*
TiC	3140	TiN	2950
HfC	3890	ZrN	2980
ZrC	3530	TaN	3090
W_2C	2860	ZrB	2990
NbC	3500		

Analytical. Zirconium is determined by precipitation as the hydrated oxide, ignited and weighed as ZrO_2. *It* can be detected by adding a drop of *certain* organic compounds (alizarin or β nitroso-α-naph-

earth's crust. Because of the very reactive nature of the metal it is found only in the combined state. The most important ore is zircon ($ZrSiO_4$) and the next is baddeleyite (ZrO_2). Zirconium is generally found associated with titanium minerals (ilmenite, rutile), and the production of zirconium has increased along with titanium production. In its ores, zirconium has 0.5–5% hafnium associated with it.



Uses. Most of the zirconium consumed is in the form of compounds for ceramic and refractory use; however, uses for the metal are becoming more numerous and important

Zircon is widely used as a refractory because of its ability to withstand high temperatures. It softens between 1600 and 1800°C and melts at 2190°C. It has low thermal expansivity and good abrasion resistance High-grade zircon sand is applied to the surface of ordinary sand molds to produce good finishes on cast parts. In the ceramics industry zirconia (ZrO_2) and zircon are used for enamels, porcelains, and glazes. When heated, zirconia emits an intense, white light and has been used as a mantle on gas flames

The metal is used in the following ways: in nuclear reactors as a structural and container material because of its low neutron-absorption cross section (when free of hafnium), its exceptional corrosion resistance, and its mechanical strength, in electronic tubes as a "getter" because of its ability to absorb large amounts of oxygen, nitrogen, and hydrogen; and in chemical equipment as containers, pumps, and valves because of its corrosion resistance. The metal has been tested for such surgical uses as skull plates, sutures, and screws and found to compare favorably with tantalum.

Production of metal. Zirconium metal is quite ductile when pure; however small concentrations of oxygen, nitrogen, and hydrogen render it brittle and impossible to fabricate. Thus the large-scale production of ductile metal was historically difficult. Two processes have been developed for commercial production.

The Kroll process consists basically of the reduction of zirconium tetrachloride vapor by liquid magnesium. The detailed process is complicated by the need to exclude oxygen and nitrogen. It consists of the following six steps: (1) Zircon is fired with carbon in an arc furnace to form zirconium carbo-nitride. The silicon is lost as silicon monoxide vapor. (2) The carbo-nitride is heated in chlorine gas to form zirconium tetrachloride. (3) The tetrachloride is purified by subliming it in an inert atmosphere. This step purifies it of oxides, which would render the final metal brittle. (4) The tetrachloride is vaporized and allowed to react with liquid magnesium. Careful control of this step is required to yield massive metal instead of powder. (5) The magnesium chloride is melted away from the zirconium under vacuum. (6) The resultant zirconium "sponge" is crushed, pressed into bars, and arc-melted under inert atmosphere into ingots. The ingots are inspected by testing their hardness, which is a measure of the embrittling impurity content. If the hardness is too great, the ingot is rejected See TITANIUM METALLURGY.

In the van Arkel, or iodide, process zirconium sponge of high oxygen content is prepared and then purified by converting it to the iodide, which is then decomposed to metal In practice the impure metal is made by the reduction of zirconia with calcium metal at high temperatures. The metal results as a fine powder in a matrix of calcium oxide, which is removed by leaching with dilute hydrochloric acid. The wet, powdered metal is loaded along the walls of a cylindrical container at whose center is a tungsten or molybdenum filament. A small amount of iodine is added and the system pumped down to a good vacuum The walls of the container are heated to about 200°C (at which temperature zirconium reacts with iodine to form the tetraiodide), and the filament heated by passage of an electric current to about 1300°C (at which temperature the tetraiodide is decomposed to the metal and iodine) The metal plates out on the filament so that the electric current through the filament must be increased continuously to maintain its temperature The net result is to transfer zirconium from the impure sponge to the filament as the pure metal, free of embrittling impurities

In both these processes the hafnium content of the zirconium metal is the same as in the original ore. For use in nuclear reactors the hafnium must be removed because hafnium absorbs neutrons about 700 times as readily as zirconium. "Hafnium free" zirconium (containing less than 0.02% hafnium) is produced by separating hafnium from the zirconium ore before reduction This can be done by ion-exchange separation of a sulfate solution, fractional distillation of halide addition compounds such as $ZrCl_4 \cdot 2POCl_3$, or fractional precipitation of the phosphate

Properties of metal. Zirconium is a lustrous, silver metal Its density is 6.49 g/cm³ It melts at 1850°C and boils at about 3580°C. Below 862°C its crystal structure is hexagonal close-packed (α form) and above that temperature it is body-centered cubic (β form). From measurements between 1675 and 1780°C the vapor pressure is given by

$$\log_{10} p = -\frac{31066}{T} - 2.115 \times 10^{-4}T + 10.2150$$

where T is in °K and p in mm of mercury. Its electrical resistivity (40 micro-ohm-cm at 0°C)

[illegible]

of neutrons is [illegible] compared to the following values for other structural materials: Fe, 2.4; Ni, 4.5; Cu, 3.5; Al, 0.22; Mg, 0.06.

The metal does not react with the common gases at room temperature and it remains bright and shiny in air indefinitely. At elevated temperatures it is very reactive to all but the inert gases, reacting in two ways, (1) to form a separate phase or compound (ZrO_2, ZrN), and (2) to dissolve large amounts of the gas to form solid solutions that still maintain the structure and some of the properties of the metal. At 700°C as much as 30 atom % oxygen, 20 atom % nitrogen, and 50 atom % hydrogen will dissolve in the metal. The metal reacts with the halogens at 200°C to form only the normal tetrahalides. Finely divided zirconium (powder or chips) is pyrophoric and hazardous to handle and store. Spontaneous explosions have been reported for both the wetted sponge and stored scrap (wet or dry).

The resistance of the metal to corrosion by aqueous solutions depends somewhat on conditions but in general is very good. Its resistance to alkalies is better than that of tantalum, titanium, or stainless steels. Its resistance to acids is very good except for hydrofluoric acid or very concentrated sulfuric or phosphoric acids. In spite of this resistance to acids and alkalies, its corrosion resistance to pure water and steam is not good. The alloy Zircalloy-2 (Zr with 1.5% Sn, 0.12% Fe, 0.05% Ni, 0.10% Cr) with good resistance to water and steam has been developed.

Compounds. Zirconium is a member of subgroup IVb of the periodic table and has a normal valence of 4. In some compounds, such as the high-melting ZrO_2, the bonds are largely ionic, whereas in others its bonds are at least partly covalent, as in the relatively low-melting halides. The degree of ionic character in the tetrachlorides of the group IVb elements increases with atomic number as evidenced by their melting points.

Tetrahalides of group IVb elements

	Melting point, °C	Boiling point, °C
$TiCl_4$	−30	136
$ZrCl_4$	437	331 (sublimes)
$HfCl_4$	432	317 (sublimes)
$ThCl_4$	820	720–750 (sublimes)

The tetrapositive zirconium ion (Zr^{4+}) does not exist in aqueous solution. Because of the high charge and small size of the ion it combines strongly with the negative ions. The simplest ion formed in solution is the zirconyl ion (ZrO^{++}). In sulfate solutions the zirconium is found in negatively charged ions, $ZrO(SO_4)_2^{--}$, and in complex chains of ZrO^{++} and SO_4^{--} links.

Zirconium tetrachloride can be prepared by the action of chlorine on the carbide and similar compounds, or on the metal. It is very hygroscopic, and it reacts violently with water to form zirconyl chloride ($ZrOCl_2$), which dissolves readily and hydrolyzes to give very acid solutions. The di- and trichloride have been made by reduction of the tetrachloride with aluminum at elevated temperatures. The bromides and iodides are very similar to the chlorides.

The sulfates of zirconium are numerous and complex. They range from the simple sulfates, $Zr(SO_4)_2$ and $Zr(SO_4)_2 \cdot 4H_2O$; through the sulfatozirconylates, $H_4ZrO(SO_4)_3$ and $H_2ZrO_2SO_4$; to the chain compounds, such as $HO(OZrSO_4)_3ZrOOH \cdot 14H_2O$.

Zirconium dioxide or zirconia has three crystal forms: the monoclinic (stable below 1000°C); the tetragonal (stable from 1000 to 1900°C); and the cubic (stable above 1900°C), which melts at 2680°C. When subjected to prolonged heating, zirconia tends to crack because of conversion from one crystal form to another. By incorporating about 5% CaO or MgO such cracking is avoided because the zirconia remains in the cubic form. This product is called stabilized zirconia.

Zirconium can absorb hydrogen to form several phases. The commercial zirconium hydride has [illegible]

On heating, it loses its hydrogen and leaves powdered zirconium. It is used in powder metallurgy and as a getter in vacuum tubes.

In certain compounds the valence of zirconium is less than four, for example, ZrN, ZrB. In these compounds and also in ZrC the structural arrangement of the zirconium atoms is essentially that of the metal, whereas the nonmetal atoms are located in sites between the zirconium atoms. The stoichiometry (that is, the ratio of number of metal atoms to nonmetal atoms) of these compounds is determined by the number of vacant sites in the zirconium structure rather than by the normal valence rules. These interstitial compounds are marked by high melting points and great hardness (8–9 on the Mohs scale). Their electrical conductivities are fairly good, and they increase with temperature, as do those of metals. The melting points of some typical interstitial compounds are given in the list.

Compound	Melting point, °C	Compound	Melting point, °C
TiC	3140	TiN	2950
HfC	3890	ZrN	2980
ZrC	3530	TaN	3090
W_2C	2860	ZrB	2990
NbC	3500		

Analytical. Zirconium is determined by precipitation as the hydrated oxide, ignited and weighed as ZrO_2. It can be detected by adding a drop of certain organic compounds (alizarin or β-nitroso-α-naph-

thol) which give highly colored precipitates of zirconyl salts. *See* CERMET; HAFNIUM; PYROMETALLURGY, NONFERROUS; REACTOR, NUCLEAR; REFRACTORY; TRANSITION ELEMENTS. [D.D.CU.]

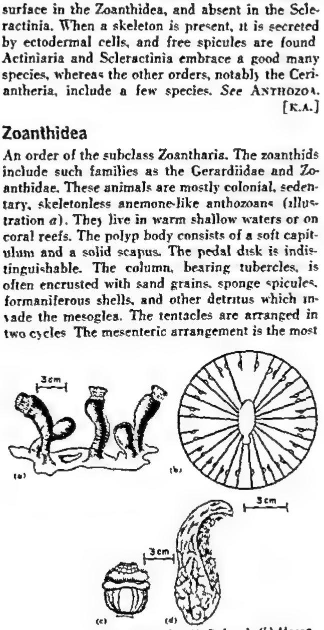
Bibliography: W. B. Blumenthal, *The Chemical Behavior of Zirconium*, 1958; G. L. Miller, *Zirconium*, 2d ed., 1957; H. Remy et al., *Treatise on Inorganic Chemistry*, 1956.

Zoantharia

A subclass of the Anthozoa. Zoantharians are monomorphic anthozoans, most of which have retractile, simple, tubular tentacles. The tentacles vary in number from six to several hundred. The mesenteries are six or some multiple of six and may be complete or incomplete in the Actiniaria and Scleractinia, while in other groups they are not constant. The siphonoglyphs are found in both edges of the stomodeum in the Actiniaria and Antipatharia, on the asulcal surface in Ceriantheria, on the sulcal surface in the Zoanthidea, and absent in the Scleractinia. When a skeleton is present, it is secreted by ectodermal cells, and free spicules are found Actiniaria and Scleractinia embrace a good many species, whereas the other orders, notably the Ceriantheria, include a few species. *See* ANTHOZOA. [K.A.]

Zoanthidea

An order of the subclass Zoantharia. The zoanthids include such families as the Gerardiidae and Zoanthidae. These animals are mostly colonial, sedentary, skeletonless anemone-like anthozoans (illustration *a*). They live in warm shallow waters or on coral reefs. The polyp body consists of a soft capitulum and a solid scapus. The pedal disk is indistinguishable. The column, bearing tubercles, is often encrusted with sand grains, sponge spicules, formaniferous shells, and other detritus which invade the mesoglea. The tentacles are arranged in two cycles The mesenteric arrangement is the most



(a) Colony of Zoanthidae (after Y. Delage). (b) Mesenteric arrangement of Zoanthidae (after Y. Delage). (c) Semper's larva, *Zoanthina tentaculata* (after Senna). (d) Semper's larva, *Zoanthella galapagoensis* (after Senna).

specific (illustration *b*). The sulcal and asulcal directive pairs consist of parts of equal size respectively. Only the former develops into complete mesenteries, the latter is always incomplete. All the other pairs consist of unequal portions. Additional mesenteries develop in the ventrolateral exocoele only. The larger mesenteries bear filaments and gonads. The musculature is poorly developed.

The life cycle is incompletely known. Two types of pelagic Semper's larvae are known (illustration *c,d*), the *Zoanthina* with a girdle of long cilia and 12 mesenteries and the *Zoanthella* with a ventral band of long cilia extending from the oral to posterior end. Daughter polyps arise by budding from the stolon or the polyp base, or by longitudinal fission. *See* ZOANTHARIA. [K.A.]

Zodiac

A band of the sky extending 8° on each side of the ecliptic, within which the Moon and principal planets remain. It is divided into 12 conventional signs, each containing 30° of celestial longitude. These signs are named Aries, Taurus, Gemini, Cancer, Leo, Virgo, Libra, Scorpio, Sagittarius, Capricornus, Aquarius, and Pisces. The names are identical with the names of the constellations through which the Sun moves, but the actual constellations are not each 30° in extent, nor does the Sun actually enter the constellation of Aries at the vernal equinox, although the vernal equinox is sometimes conventionally called the first point of Aries. The signs of the zodiac are without practical astronomical significance [C.M.C.]

Zodiacal light

A diffuse band of luminosity occasionally visible on the ecliptic. It is sunlight diffracted and reflected by dust particles in the solar system within and beyond the orbit of Earth. Zodiacal light is best seen after evening twilight or before morning twilight when the ecliptic is at right angles to the horizon. In the Northern Hemisphere this is in the evening during springtime, and in the morning during autumn

The light steadily increases from about 170–30° away from the Sun

Fraunhofer corona. Sunlight scattered within Earth's atmosphere normally prevents observation of zodiacal light at elongations (angular distance from the Sun) of less than 30° During a solar eclipse, however, the scattered light is reduced and, at positions near the Sun, the zodiacal light may be intense enough to observe.

Part of solar coronal light is due to diffraction by zodiacal particles. Superposed on the diffracted light, is light scattered by electrons in the solar corona itself. The Doppler effect of the rapidly moving electrons blurs the solar Fraunhofer absorption lines, leaving a continuum that dilutes the Fraunhofer lines in the diffracted light. This dilution allows the two components of the corona to be separated. The intensity due to the zodiacal particles is well represented by an extrapolation of

the zodiacal light observed in elongations between 170 and 30°. A constant space density of particles, at distances greater than 0.1 astronomical units (AU) is sufficient to explain the observations (Presumably such matter is vaporized by the sun at lesser distances.)

Gegenschein. At elongations of 180°, zodiacal light shows a marked increase in brightness. This patch of light, ordinarily about 10° in diameter, is called the Gegenschein or counterglow. It is attributed to the increased efficiency of reflections at large phase angles by diffusely reflecting particles.

Poynting-Robertson effect. A relativistic study of the equations of motion for a body moving in a radiation field was made by H. P. Robertson following an approximate solution of this motion problem by J. H. Poynting. If absorption and isotropic reemission of radiation occur, the body experiences a resisting force proportional to its velocity. For a body in orbit about the sun, the semimajor axis must decrease for the orbit to be commensurate with orbital energy. Eventually, then, the body spirals into the Sun. Because the radiant energy absorbed is proportional to the area of the body and because the resisting force is inversely proportional to the mass, small bodies are influenced more than large. H. C. van der Hulst showed that the zodiacal particles are generally smaller than 0.35 millimeters in radius. Particles in this size range and with relatively small initial orbits (1–10 AU) are thrown into the Sun within a few million years and at a rate of about 1 ton per second. F. L. Whipple has estimated that meteoric material is introduced into the solar system by cometary disintegration at a rate of about 30 tons per second. This probably represents the source of supply for the zodiacal particles.

[R.E.MC.]

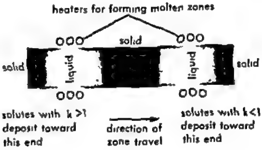
Zone refining

A technique for producing high purity materials. It is used to alter the distribution of soluble impurities that exhibit a difference in solubility between the liquid and the solid phases of the base material. In zone refining a relatively narrow molten zone is passed through an otherwise solid ingot. The zone width is small in comparison to the ingot length, and the rate of zone travel is slow.

Zone refining is used primarily to produce semiconductor materials and metals of high purity, in some cases containing only a few parts per billion of impurities. Much less effort is required to attain this high degree of purification by zone refining than by the older technique of fractional crystallization. For a discussion of materials purification by solute control, see CRYSTALLIZATION, SEMICONDUCTOR.

The method is also used to distribute solute elements evenly in a material in order to produce a homogeneous end product, or to produce discontinuous distributions of solutes in materials for use in transistor or diode applications. Additional uses are for joining, single crystal growing, and diffusivity measurements.

In practice, a number of molten zones are passed slowly and in one direction through an ingot by some means, such as by moving heating coils over a stationary ingot.



Zone refining

The primary factor in determining how the solute redistributes during zone melting is the distribution coefficient, k, which is the ratio of the [illegible] the liq- [illegible] with dis- [illegible] ore soluble in the liquid than in the solid; therefore, they are swept along the ingot in the molten zone and [illegible] to be rejected by the liquid and deposited at the starting end. Solute elements with distribution coefficients of one are not redistributed since they have the same solubility in the liquid as in the solid. After the refining process, the ends of the ingot containing the bulk of the impurities are cropped.

In the process known as zone leveling, the molten zones are moved back and forth in the ingot, producing a uniform or homogeneous material. The ends are cropped to remove those portions with uneven solute distributions. See METALLURGY.

[R.L.SM.]

Bibliography: W. G. Pfann, *Zone Melting*, 1958; A. Weissberger, (ed.), *Technique of Organic Chemistry*, vol. 3, 1950.

Zoogeographic region

A major unit of the earth's surface characterized by fauna[illegible] concept [illegible] animals [illegible] ing earth [illegible] in their vagilities and tolerances to environmental conditions, no two major groups of animals display complete coincidence in their geographic limits. As a result, delimitation of the several regions is difficult and such patterns as have been defined are not universally applicable to the entire animal kingdom.

Fortunately, insofar as existing animals are concerned, there appears to be a rough average geographic pattern. This pattern is best mirrored in the distribution of birds and mammals. Considerable coincidence in the geographic patterns of these

two groups has possibly resulted from the fact that they arose and dispersed relatively recently in geologic time and, as a result, their extant distributions have been controlled by major geographic barriers of the not too distant past. These patterns reflect to a lesser degree the distributions of other major groups. It appears, therefore, that several parts of the earth's surface have, through isolation, served as major centers of evolution, and in the final analysis, it is these centers that are recognized as zoogeographic regions. Broad zones of transition have developed between the major zoogeographic regions owing either to the ability of some animals to transcend major barriers or to the disappearance of the barriers. It is the occurrence of such transitional areas that has led to most of the difficulties in the plotting of zoogeographic boundaries.

Classification. Although earlier attempts had been made to partition the earth into zoogeographic regions, it was in 1858 that P. Sclater, on the basis of bird distribution, presented the first practical regional classification. This was extended by A. Wallace in 1876 and has since undergone some further modification, but the pattern described by Sclater has remained basically sound. The table correlates the classical Sclater-Wallace system with K. Schmidt's 1954 review of the problem.

Zoogeographic regions

Schmidt (1954)			Classical system regions
Realms	Regions	Subregions	
Arctogaean	Holarctic	Arctic	Nearctic-Palearctic
		Nearctic, Caribbean	Nearctic
		Palearctic	Palearctic
	Paleotropical	Oriental	Oriental
		Ethiopian, Malagasy	Ethiopian
Neogaean	Neotropical	Neotropical	Neotropical
Notogaean	Australian	Australian, Papuan, New Zealandian	Australian
	Oceanian	Oceanic	
		Antarctic	Unclassified

Although the Palearctic and Nearctic regions share many faunal elements, particularly insofar as eastern Asia and eastern North America are concerned, they do display important faunal differences. The tropical elements of the two regions, especially, show considerable divergence. That of Palearctica has been derived from the Ethiopian and Oriental regions, whereas the Nearctic tropical element has stemmed, for the most part, from Neotropica. Thus, although the two northern regions are frequently combined as Holarctica in order to express their many resemblances, they are more generally treated individually.

Palearctic region. This region includes all Eurasia north of the Tropic of Cancer, excepting parts of the Arabian Peninsula and northwestern India. Both the East (China) and the West (Europe) originally supported deciduous forest. To the North, the boreal forest and the tundra are circumpolar. The vast interior of Asia is largely grasslands with considerable desert on the high plateaus.

The Palearctic fauna is difficult to characterize. As a generalization, it is composed of three elements. One is world-wide or almost so, the second is more or less circumpolar, and the third has been derived from the Old World tropics. At the family and subfamily levels there are few exclusives. Furthermore, the extensive longitudinal spread of the region that leaves the temperate forests of East and West disjunct, because of the subhumid lands of the interior, produces a number of east-west faunal differences.

Fishes. Except for Australia, the fish fauna is dominated by the world-wide cyprinids or minnows. With North America, the Palearctic shares the percids or perches, paddle fish, and suckers. The last two are restricted to the Asian area of the Palearctic. The cobitids (loaches) and a few catfishes are shared with the Old World tropics.

Amphibians. Among the amphibians the hynobiid salamanders are endemic to eastern Asia. Newts are Holarctic in their distribution. The plethodontid salamander of Europe is a single representative of a large, New World group, whereas, in contrast, the hellbender of China is related to that of eastern North America. The tailless amphibians are largely the widely distributed bufonids or toads, and the hylid and ranid frogs.

Reptiles. The reptilian fauna is poor. The emydid turtles and the varanid and agamid lizards have all been derived from the South. Among the snakes, the majority belong to the harmless and almost world-wide colubrids. True vipers in the West have been derived from the South and the pit vipers of the East are shared with the New World.

Birds. The Palearctic avifauna reflects the general faunal picture outlined above. A more or less world-wide element is represented by the hawks, woodpeckers, swallows, and finches. Circumpolar groups include the grouse, waxwings, and creepers. Tropical elements, largely migratory, include the Old World flycatchers, larks, and starlings. Such Oriental derivatives as pheasants, cuckoos, shrikes, and white-eyes occur only in eastern Asia. Only a single family, the hedge-sparrows, is endemic to the region.

Mammals. The mammalian fauna resembles the avifauna in its geographic relationships. Except for Australia, world-wide groups include rabbits, cricetid mice and rats, squirrels, and cats. Circumpolar types include the beaver, jumping mice, and the pikas. The murids (Old World mice and rats), the dormice (also African), and the panda (restricted to eastern Asia) are essentially Old World groups.

Nearctica. Nearctica includes all of North America north of the edge of the Mexican Plateau. The vegetation pattern is similar to that of Palearctica. Across the North, the boreal forest and the tundra are both transcontinental. The entire East and the coastal fringe of the West support temperate forest cover. The central region, although bisected by the

forested Rocky Mountains, is largely grasslands and deserts.

In general, the faunal picture is also similar to that of Palearctica. There is considerable east-west diversity and a mixture of world-wide, tropical, mostly from South America, and holarctic groups. Endemism is however, better marked.

Fishes The fish fauna is richest east of the Mississippi River. World-wide cyprinids are abundant. Holarctic or at least Nearctic and Asian groups include the suckers, paddlefishes, and perches. A few South American elements, notably characins and cichlids, extend up to the Rio Grande River. Exclusive groups are represented by the bowfin, mooneyes trout, and basses.

Amphibians. Among the amphibians, the salamanders, especially the plethodontids, are particu-

ranids. A rather interesting, primitive frog, *Ascaphus*, is known only from the high mountains of the west coast. It belongs to a group that is otherwise restricted to a small island off the coast of New Zealand.

Reptiles The reptile fauna includes such almost world wide groups as the gekkonid and scincid lizards, and the colubrid snakes. It shares with Eurasia the emydid turtles, the anguid lizards and the pit vipers. More restricted in their distributions (*mainly in* South America) are the coral snakes, the teiid lizards, the iguanid lizards (also on Fiji and in Madagascar), and the peculiar, poisonous gila monster.

Birds For the most part, the avifauna comprises world wide and holarctic groups and a number of migratory neotropical elements. Ducks, pigeons, hawks, kingfishers, swallows, and finches are all widely distributed. Holarctic groups include grouse, creepers, and waxwings. Strictly New World elements are represented by the New World flycatchers, vireos, orioles, and hummingbirds. The turkey is almost endemic, extending as far south as northern Central America.

Mammals Of the mammals, the cats, bovids, rabbits, cricetid mice, and squirrels are wide-ranging. Mammals more typical of Holarctica include the jumping mice, microtine mice, and the beaver. Two exclusives occur in the region, the pronghorn of the western deserts and the mountain beaver of the Northwest.

Ethiopian region. In the modern parlance, this region refers to Africa south of about the mid-Sahara. Some zoogeographers also include the Arabian Peninsula, which with Mediterranean Africa is considered by others to be a transition province. The region is entirely tropical, except for the subtropics of the extreme south, although temperatures are moderated by the continent's upland nature. Tropical forests fringe the Gulf of Guinea and extend inland through the Congo basin and occur also in the eastern mountains. Deserts include the Sahara, the Kalahari, and the Somaliland coast. Most of the remainder of the continent supports scrub forest and grasslands.

Faunally, the Ethiopian region is most closely related to the Oriental region. Nevertheless it shares a number of groups with Holarctica, whereas South American relationships are evident in its fish and turtle faunas.

Fishes Among the fishes, the widely distributed cyprinids are abundant. The lung fish has South American affinities, as well as a wealth of characins and the cichlids which have undergone extensive differentiation in the East African lakes. The primitive birchirs as well as several groups of primitive bony fishes are exclusives.

Amphibians. The amphibian fauna is not particularly remarkable. It includes the widely distributed caecilians, bufonids, and ranids. It shares the primitive pipids with South America. The true tree frogs are lacking, but their niche is filled by the rhacophorids which are also Oriental; some authorities consider the African representatives of this group to be a separate family restricted to the continent. A family of frogs related to the world-wide narrow-mouth frogs is exclusively African.

Reptiles Although it includes such widely distributed groups as the skinks among the lizards and many harmless colubrid snakes, the reptile fauna is essentially pantropical. Notable examples of these are the sideneck turtles, crocodiles, gekkos, and the worm snakes. Oriental affinities are evident in the chameleons, an egg-eating snake, the cobras, and the true vipers.

Birds The avifauna, although large, is represented by few endemics above the generic level. The bulk of the avifauna is composed of widely distributed groups such as the owls and hawks, kingfishers, swallows, and true finches. The region has much in common with both Europe (Old World flycatchers starlings, and orioles) and the Orient (broadbills, hornbills, and honey-guides). Exclusives include the ostrich, the secretary bird, certain genera of guinea fowl, widow-birds and tick-birds.

Mammals The spectacular big game, such as the elephant, rhinoceros, hippopotamus, a variety of antelopes, horses (zebras), cats, and giraffes, hardly serve to bring out the true nature of the mammalian fauna. The region supports such widely distributed groups as mustelids, rabbits, and many rodents. Oriental affinities are indicated by rhinoceroses, Old World monkeys, great apes, scaly anteaters, and bamboo rats. Africa also possesses a number of endemic or near-endemic mammals of which the golden mole, elephant shrews, hyraxes, and the giraffe are best known.

The fauna of Madagascar and the islands of the Indian Ocean presents a number of interesting problems that are not likely to be settled in the immediate future. Madagascar supports a fauna that has much in common with Africa—side-neck turtles, rhacophorid frogs, chameleons, and many birds and bats. In contrast, many mainland groups are absent, such as primary fresh-water fishes, pythons and cobras, the ostrich and tick-birds, hystri-

comorph rodents, Old World monkeys, and great apes. The island possesses a number of endemics including a group of narrow-mouth frogs, some rollers and vangas among the birds, and a variety of lemurs. A few Oriental relationships are also apparent.

The Seychelles and Mascarene Island groups of the Indian Ocean support small but interesting faunas. The flightless birds of the Mascarenes, especially the now extinct dodo, are particularly noteworthy.

Oriental region. This region encompasses tropical Asia from the Iranian Peninsula eastward through the East Indies to and including Borneo and the Philippines. Its exact boundaries, however, are difficult to define because of broad areas of transition between it and adjacent regions. Aside from the Thar desert and isolated areas of semidesert, especially along the east coast of India, the region originally supported forest growth.

The region appears to have served both as a center of dispersal for cold-blooded vertebrates and as a crossroads through which various other groups have passed. Although its fauna bears much in common with the Ethiopian region, it shows both Palearctic and Australian affinities.

Fishes. Although ancient groups of fishes that [illegible] group, the loaches, is almost exclusively Oriental. Several families of catfishes are endemic to it and still others of this same group are shared with Africa. The climbing perch make up another distinctive group and further emphasize African affinities.

Amphibians. The amphibian fauna is large. Caecilians are present although the salamanders barely enter the region. Among the tailless amphibians are the widely distributed bufonids, ranids, and narrow-mouth frogs. It has many rhacophorid frogs which further attest to African faunal affinities.

Reptiles. The reptilian fauna is largely shared with other regions. Widely distributed groups include the skinks and harmless colubrid snakes, whereas such pantropical groups as the gekkos, worm snakes, and pythons are all well represented. [illegible] and the co- [illegible] and true [illegible] the Old World. The pit vipers are shared with the New World. The Oriental region also supports several small endemic families of snakes and the surrounding seas are well populated with sea snakes.

Birds. The avifauna, like that of Africa with which it shares many groups, is poor in endemics above the generic level. Widely distributed groups such as woodpeckers, pigeons, and jays comprise the bulk of the fauna and pheasants are particularly characteristic of the region. A few Australian groups such as the frogmouths and wood swallow enter the Orient from the Southeast. The region boasts only a single exclusive family, the fairy bluebirds.

Mammals. The mammal fauna, although represented by such wide-ranging groups as weasels, rabbits, and squirrels, includes a number of items shared with Africa as well as a good representation of endemics. African affinities are evident in Old World monkeys, scaly anteaters, the rhinoceros, elephant, and the fruit bats. Among the endemics may be mentioned the flying lemurs, tree shrews, and tarsiers.

The island archipelago to the east and southeast, which includes Sumatra, Java, and the Philippines, supports a vertebrate fauna that is very definitely Oriental in character. As is generally true of archipelagos, the geographic patterns of various animal groups form a complex mosaic that is accompanied by depauperization. This last is particularly evident in the northernmost islands of the Philippines.

Neotropical region. This area includes Mexico south of the Mexican Plateau, the West Indies, Central America, and South America. The first three are frequently treated as a transition zone between Nearctica and Neotropica.

Although about two-thirds of the region lies within the tropics, its plateau and montane character is responsible for considerable areas with nontropical temperatures. Generally speaking, the Amazon Basin and the east coasts support high rainforests. Scrub forests and grasslands clothe much of the interior of South America, whereas desert occupies much of the continent's west coast and Patagonia. The Andes and the Central American mountain systems provide a variety of vegetation belts.

Owing to a history of isolation through much of the Cenozoic Era, the South American fauna is composed of two very distinct elements. One is of considerable age with some pantropical groups. The other is a younger element that has invaded the region relatively recently from North America.

Fishes. The fish fauna, although represented by but few major groups, is, nevertheless, very rich and with many endemics at the family level. Representatives of more ancient types include the pantropical osteoglossids and a lungfish which is shared with Africa. Other African affinities are evidenced by the characins, cichlids, and nandids. The widely distributed cyprinids are absent and a number of other northern types barely enter Central America. Among the more notable endemics are the gymnotid (electric) eels and many families of catfishes.

Amphibians. The amphibian fauna includes a wealth of the widely distributed hylids and leptodactylids (these are especially abundant in Australia) and such pantropical or near-pantropical groups as the caecilians and brachycephalid frogs. With Africa it shares the pipid frogs and with the Orient the narrow-mouth frogs. The wide-ranging ranids and salamanders are poorly represented. The region supports but a single exclusive frog family, the primitive *Rhinophrynus* which is restricted to southern Mexico.

Reptiles. The reptile fauna includes skinks and harmless colubrid snakes among the near-cosmo-

politan groups. Pantropical representatives include the side-neck turtles, some related to those of Africa and others to those of Australia, gekkos, worm snakes, and the coral snake family. It shares the pit vipers with Nearctica and the Orient. Other groups confined to the New World include the many iguanid lizards, several on Madagascar and one on Fiji, and the teiids. Endemic reptiles include several families of turtles that are restricted to northern Central America, the caimans, and the boas.

Birds. The rich avifauna has led to the designation of South America as the bird continent. Although many widely distributed groups such as herons, ducks, hawks, parrots, trogons, and thrushes are well represented, about 50% of the Neotropical families are endemic. About one-third of the bird species belong to the exclusive furnaroid group, ant birds, ovenbirds, and woodhewers, whereas the wood warblers, hummingbirds, and flycatchers, although shared to a greater or lesser extent with North America, comprise most of the remaining endemic species. Other exclusives include the flightless rheas, tinnamous, toucans, and cotingas.

Mammals. The mammal fauna further emphasizes South America's history of isolation. The list of endemics is long and includes several groups of marsupials, sloths, New World monkeys, and most of the hystricomorph families. With the Old World it shares the camels, llama, and the tapir, and with North America the armadillos, peccaries, and the pocket gophers. More wide-ranging groups include weasels, cats, squirrels, cricetid mice and rats, and some bats.

Central America and lowland Mexico is an area of faunal overlap between the Nearctic and Neotropical elements. The region is a pathway of considerable environmental diversity and it has been utilized by northern groups dispersing southward and by southern groups dispersing northward, both to varying degrees. As a result a complex zoogeographic mosaic, still largely unstudied, obtains throughout the region.

The West Indies support a depauperate fauna of both Nearctic and Neotropical affinities. The representation of the various groups through the archipelago appears to be in direct proportion to their ability to cross water barriers. Although many schemes have been presented to explain the populating of the islands, most evidence indicates that most groups were transported across water barriers from Central America.

Australian region. Included in this region are continental Australia, New Guinea, Tasmania, and lesser islands through the Solomon group. The boundary between it and the Oriental region has long been debated. It was originally placed by Wallace (Wallace's Line) well to the West between Bali and Lombok, between the Celebes and Borneo, and south of the Philippines. The easternmost boundary now generally accepted lies just west of Aru and New Guinea. Modern zoogeographers recognize that the fauna of the archipelago between the two extremes is transitional. Schmidt refers to it as the Celebesian transition subregion, the Wallacea of many authors.

Continental Australia possesses both tropical and subtropical climates. Probably no less than 75% of its area supports desert or grassland cover. Forests and forested grasslands are restricted to the northern, eastern, and southwestern fringes of the continent. Eucalyptus forest, almost endemic to Australia, predominates in the southeast. New Guinea has a cover of considerable rainforest and wet mountain forest.

As a result of its long history of isolation, the region has a fauna that includes some very ancient endemic groups. These have survived here, and in South America, and have undergone considerable adaptive radiation. In addition Australia also possesses a more recent element of Oriental affinities.

Fishes. The fish fauna is poor. Aside from an osteoglossid, an ancient group shared with South America, Africa, and southeastern Asia, and an exclusive family of lungfish, only distantly related to that of South America and Africa, the fishes of Australia are salt-tolerant and of wide distributions.

Amphibians. Of the amphibians, only the frogs are represented and these by only four families. In Australia proper the leptodactylids and hylids, both almost cosmopolitan, comprise about 90% of the amphibian fauna. New Guinea, in contrast, supports a wealth of the narrow-mouth frogs.

Reptiles. Reptiles are well represented throughout the region. Pantropical groups include crocodiles, side-neck turtles, and gekkos, whereas skinks and colubrid snakes are even more wide-ranging. General Old World affinities are expressed by the agamid and varanid lizards and the pythons. Among the poisonous snakes the region possesses only the elapids. A single reptilian family, the pygopod lizards, are endemic.

Birds. The avifauna, although lacking many widely distributed families such as pheasants, woodpeckers, and finches, is represented by such near-cosmopolitan groups as the pigeons, kingfishers, and parrots. Endemics include the spectacular emu and cassowary (both flightless), the bird of paradise (especially characteristic of New Guinea), and the strictly Australian scrubbirds and lyrebirds.

Mammals. The mammalian fauna of the region is spectacular in the adaptive radiation displayed by the marsupials. Although not exclusive, they comprise the bulk of the mammal fauna. These and the monotremes testify to the long-continued isolation of Australia proper. Bats and the murid mice and rats are the only other mammalian groups native to the region. The dingo (wild dog) and the rabbit are introductions.

To the east of Australia and New Guinea, the [illegible]

fresh-water fishes, turtles, snakes, and mammals, aside from bats and a few introduced species such as the deer. The islands have served as a refuge for such ancient groups as the lizardlike tua-

tara *Sphenodon*, a tailed frog which is shared with northwestern North America, and the moas and kiwis.

Toward the northeast strictly fresh-water fishes do not extend beyond New Guinea, terrestrial mammals and frogs reach only to the Solomons, snakes to Fiji, and lizards fall just short of Samoa. The Hawaiian Islands, aside from introduced forms, include among their vertebrate fauna only a single bat and a variety of endemic birds of New World affinities. *See* BIOGEOGRAPHY; BIOTIC ISOLATION; ISLAND FAUNAS AND FLORAS; PLANT GEOGRAPHY. [L.C.S.]

Bibliography: P. J. Darlington, *Zoogeography*, 1957; Karl P. Schmidt, Faunal realms, regions, and provinces, *Quart. Rev. Biol.*, 29:(4), 322–331, 1954; P. L. Sclater, On the general geographical distribution of the members of the class Aves, *J. Proc. Linnaean Soc.* (*Zool*), 2:130–145, 1858; A. R. Wallace, *The Geographical Distribution of Animals*, 2 vols. 1876

Zoogeography

The science that attempts to describe and explain the distribution of animals through space and time. To accomplish this, zoogeographers inventory and analyze two sets of factors, those intrinsic to the organism (genetic), and those extrinsic to the organism (environmental). Thus the data of zoogeography are drawn from such varied fields as animal morphology, physiology and systematics, botany, paleontology, physical geography, and stratigraphy. Zoogeography, therefore, is frequently viewed as a borderline science. Neverthe-... ial field of study

... oogeography are ... study of animal distribution is by fiat zoogeography. During the twentieth century, however, investigations into a number of special phases of animal distribution have given rise to fields of study which have attained the status of sciences in their own rights. Among these may be mentioned especially ecology and limnology. The dividing line between zoogeography and ecology, for example, is frequently merely one of scale. Zoogeography operates at the global and continental levels and deals essentially with higher systematic categories (genus and above). In contrast, ecology is concerned with more local areas and trivial systematic catagories (species, subspecies, or single populations thereof). *See* ECOLOGY; LIMNOLOGY.

Objectives. The aims of zoogeography vary depending upon the interests of the investigator. One individual may seek to partition the earth into zoogeographic regions; another may essay an explanation of a geographic pattern displayed by some particular group of animals; a third may attempt to show how some geographic pattern is sustained in terms of faunal adaptations. Zoogeography, then, may have many aims, each characterized by a particular approach. These approaches have been examined by W. Allee and K. Schmidt, and viewed collectively they express the essential aims of the subject. Though not in complete accord with Allee and Schmidt, the following will serve to summarize the nature of the ends that modern zoogeographers hope to attain through the employment of the several approaches to the subject.

Systematic approach. The systematic approach takes as its point of departure a systematic group, generally some unit between genus and class. The problem is to delimit such a group, to analyze its make-up, and to describe its distribution. The assembling of these data is the chore of the systematist, and museum collections provide him with his essential materials. The resultant data may be interpreted in several ways. The investigator may undertake an explanation of the distribution of the ...

... ponents and near relatives (historical interpretation). Or, in contrast, he may analyze the distribution in terms of the morphological or physiological qualities or both that have permitted the group to radiate adaptively into the several environments that may be circumscribed by its range (ecological interpretation). D. Amadon's monograph of the Hawaiian honeycreepers is a superb example of a systematic study interpreted along both historical and ecological lines.

It was the systematic approach that engendered one of the most violent quarrels ever witnessed by zoogeography. Zoogeographers early recognized that certain more specialized groups of animals, like placental mammals, though occurring on the southern continents, show continuity in their distributions through boreal regions. In contrast many more primitive groups such as the lungfishes and side-neck turtles occur only on the southern continents. To explain discontinuities in the distributions of these more primitive groups, some zoogeographers were led to postulate ancient, transoceanic land bridges connecting the southern continents. These postulates were contrary to geological facts and were not acceptable to more conservative zoogeographers. The problem was finally resolved by W. D. Matthew who showed that the vertebrates, at least, had evolved in the north, dispersed southward and, in the case of many primitive groups, had been exterminated in the north. This hypothesis, borne out by the paleontological record, is now generally accepted.

Regional approach. The regional approach seeks to partition the earth into regions of various scales on the basis of animal resemblances. In general there are two types of zoogeographic regions, faunal and ecological. A faunal region is characterized by an assemblage of animals whose ranges show a high degree of accordance, in which endemism is generally considerable, and from which certain groups of animals are frequently absent. An ecological region is characterized by an assemblage of animals displaying homologous or

analogous adaptive responses to a particular set of environmental conditions.

Throughout most of the last half of the nineteenth century students of zoogeography devoted most of their energies to the partitioning of the globe into faunal regions (see illustration). At the global level the initial practical divisioning of the earth followed P. L. Sclater's system. Quibbling over the comparative ranks of regions, values of indicator groups and the positions of boundaries followed These investigations have been summarized by Schmidt (*see* ZOOGEOGRAPHIC REGION). At the continental level North America in particu-

ogeographic regions of the world (From K P ...hmidt, *Faunal realms, regions, and provinces, Quart ... Biol*, 29(4) 322–331, 1954)

...r was subjected to the *scrutiny of* zoogeographers ...eking a basic pattern of animal distribution ...arly attempts between 1870 and 1890 culminated ... C. H. Merriam's *life zone* hypothesis. Though ...is scheme served as a *useful* statement of animal ...stribution during the first quarter of the twentieth ...entury, it has since been largely abandoned *in* ...vor of L. R. Dice's biotic province concept. *See* MERRIAM'S LIFE ZONES

Efforts to divide the globe into ecological regions ...ave been based more upon the distribution of ...nvironment types than upon the distribution of ...nimal assemblages. Thus the present day *ecologi*...cal regions have been defined largely *in terms of* ...vegetation which seems to *summate* the effects of ...uch environmental features as climate, soils, and topography. A climax vegetation-type (forest, grassland, desert) with its associated fauna constitutes a biome or biome-type which is the zoogeographer's term for an ecological region. The fauna within any biome, through homologous or analogous structures and physiological processes, is adapted to the prevailing environmental features. In the desert biome-type, for example, these features include scarcity of water, desiccating winds, extreme diurnal temperature ranges, and lack of cover. Regardless of the locales of deserts, the faunas of these regions are adapted in similar ways to cope with the *environmental conditions* encountered therein. The fact that phylogenetically the faunas of the Great Australian Desert and of the Colorado Desert have nothing in common is beside the point. Animal life in the two regions responds in similar ways to the desert environment *See* BIOME; CLIMAX PLANT FORMATIONS

Thus whereas a faunal region is a continuous areal unit, an ecological region lacks areal continuity and may occur in widely separated parts of the earth Furthermore, a single faunal region *may* encompass several ecological regions. The Neotropical realm, for example, includes such diverse ecological regions as the Amazonian rainforest, the Guiana grasslands and the Patagonian Desert.

Explanations as to how ecologial regions are sustained as entities depend largely on the environmental morphologist and environmental physiologist Data on ecological regions have been summarized by R Hesse and expanded by Allee and Schmidt.

Faunal approach. The faunal approach essays an analysis of animal distributions in *terms of faunal groups* which have homologous histories in contrast

classic analysis of the reptile and amphibian faunas of the Americas by E Dunn. Development of this approach had to await the accumulation of data on specific groups by the systematist As more and more groups of animals were analyzed systematically and historically, pattern-types began to take form It was discovered, for example, that the reptile and amphibian fauna of Central America, nominally a part of the Neotropical realm or faunal region, is historically not homogeneous, but, rather, is composed of at least three *distinct faunal elements* These have moved into Central *America*, and sometimes through it at different times Thus *the mere reference* of the Central American fauna *to the Neotropical* realm or a region thereof indicates very little A faunal analysis provides an understanding of the composition of that fauna and how it *developed*

Because analyses of this sort present data which *are of value* to other sciences, especially as supporting evidence to the paleogeographer, the faunal approach has proven most useful.

It should be understood that in modern zoogeographical studies the several approaches are

rarely utilized independently. Most frequently the investigator weaves the several approaches together in an effort to determine the nature of a zoogeographic pattern, how it developed, and how it is sustained. [L.C.S.]

Bibliography: W. C. Allee and K. P. Schmidt, *Ecological Animal Geography*, 1951; D. Amadon, The Hawaiian honeycreeper (Aves, Drepaniidae), *Bull. Am. Museum Nat. Hist.*, 95(4), 1950; P. J. Darlington, *Zoogeography*, 1957; L. R. Dice, *The Biotic Provinces of North America*, 1943; E. R. Dunn, The herpetological fauna of the Americas, *Copeia*, 3:106–119, 1931; H Gadow, *The Wanderings of Animals*, 1913; P. E. James and C. F. Jones (eds.), *American Geography, Inventory and Prospect*, 1954; W. D. Matthew, Climate and evolution, *Ann. N.Y. Acad. Sci*. 24:1–323, 1915; C. H. Merriam, Laws of temperature control of the geographic distribution of terrestrial animals and plants, *Natl. Geograph. Mag.*, 6:229–238, 1894; A. E. Ortmann, The geographical distribution of fresh-water decapods and its bearing upon ancient geography, *Proc. Am. Phil. Soc.*, 41:267–400, 1902; P. L. Sclater, On the general geographical distribution of the members of the class Aves, *Proc. Linnaean Soc. London*, 130–145, 1858

Zoological nomenclature

The system of names used to designate the various kinds of organisms which comprise the animal kingdom. First formalized by the eighteenth-century Swedish naturalist Linnaeus, the system has been expanded and modified into a universal procedure for the naming of animal species and their arrangement in a hierarchy of appropriately designated categories of classification (taxa). Until the adoption of this method, progress in the systematic study of animals was greatly hampered and the development of a natural classification scarcely possible.

Binomial nomenclature. In the Linnaean system of binomial nomenclature, the species has a dual designation, the first element of the name representing the genus (for example, *Canis*, the various dogs), the second, a particular species in that genus (as *lupus*, the wolf) These two elements comprise the species name, which must be unique; that is, the particular combination of generic and specific name cannot be used to designate any other species of animal. It is followed by the name of the author. Thus, the correct scientific name for the wolf is *Canis lupus* Linnaeus. When, through inadvertance, or through reclassification two or more such combinations come into existance they become homonyms and the more recent are rejected Likewise, each species can have only one valid name, the first properly proposed Subsequent names become synonyms and are rejected (principle of priority). Since Greek and Latin were the common languages of educated men of the eighteenth century, it was natural to make use of these languages in forming scientific names, the generic names being selected largely from latinized Greek or from combinations of Greek words, the specific names mainly from Latin. Subsequently it has become customary to cite the name of the author following the binomial specific name, alone if the original combination of generic and specific names is still used, in parentheses if the combination has been changed. However, in order to be available for use a newly proposed name must be accompanied by a definition or description of the taxon to which it applies. A name proposed without such a definition is a nomen nudum, one with a definition which cannot be interpreted satisfactorily is a nomen dubium

Higher taxa. Names of genera and other higher categories (taxa) in classification are single words (uninomial), and must be unique, not used for other groups of animals. The generic name is singular; those of higher categories are always plural Names for categories above the genus and below the order group are based upon the root of the name of one of the included genera (the type genus) and their ending indicates the category, as "-inae," subfamily; "-idae," family. Above the family group of names, designations for higher categories, as orders, classes and phyla, do not have uniform endings, although in practice there is considerable uniformity in certain groups such as in the Insecta where the ending "-ptera" usually denotes an order.

International code. Although Linnaeus proposed an elaborate series of rules governing the naming of organisms, in subsequent years these became variously modified in practice or altered to meet the requirements of expanding knowledge. As varying procedures were introduced into practice, it became imperative that steps be taken to secure a more stable and universal nomenclature if the basic system of Linnaeus was to survive. To this end, the International Congress of Zoology, toward the beginning of the twentieth century, established an *International Code of Zoological Nomenclature* (*Règles Internationales de la Nomenclature Zoologique*) and a permanent International Commission on Zoological Nomenclature to enforce and interpret it. This code consisted of 41 articles and 20 recommendations, dealing with family, generic, specific and subspecific names, their validity, formation, and orthography, the designation of types for species and genera and related matters Of particular importance was the adoption of the principle of priority of publication, the separation of zoological nomenclature from botanical nomenclature and the selection of the tenth edition of Linnaeus' *Systema Naturae* (1758) as the starting point of zoological nomenclature This was an important international achievement. However, the retroactive application of the rules to the zoological publications of nearly 150 years created many problems and jeopardized the stability of many names then widely utilized in the fields of medicine, public health, and agriculture As a result, the Commission was subsequently provided with plenary powers to set aside the application of the

rules in certain cases, an authorization which has been broadened by recent congresses, under the *principle* of conservation of well established names (nomina conservanda) Names acted upon in this manner are placed on official lists and are not subject to further change on nomenclatural grounds.

Decisions of the International Commission have been issued in the form of Opinions, approximately 200 of which were published, prior to about 1935 by the Smithsonian Institution, Washington, D.C., subsequently by the International Trust for Zoological Nomenclature, London. By 1948, so many defects had become apparent in the international rules, so many interpretations remained unincorporated and so many new problems had become evident that a complete redraft was authorized by the International Congress of Zoology meeting in Paris Further alterations were made at Copenhagen in 1953, and at London in 1958 where a new code was drafted. *See* PLANT CLASSIFICATION

[E.C.LI]

Bibliography. E Mayr, E. G. Linsley, and R. L. Usinger, *Methods and Principles of Systematic Zoology*, 1953; E T Schenk and J. H McMasters, *Procedure in Taxonomy*, 3d ed , 1956

Zoology

The science that deals with knowledge of animal life Together with botany, the science of plants, it forms biology, the science of living things. With the great growth of information about animals, zoology has been much subdivided Some major fields are anatomy, which deals with gross and microscopic structure, physiology, with living processes in animals, embryology, with development of new individuals, genetics, with heredity and variation, parasitology, with animals living in or on others, natural history, with life and behavior in nature, ecology, with the relation of animals to their environments, evolution, with the origin and differentiation of animal life, and taxonomy, with the classification of animals *See* BIOLOGY, BOTANY

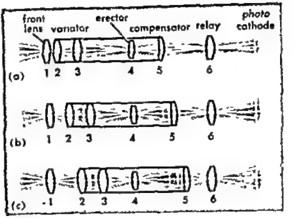
[T.L.S.]

Zoom lens

A system of lenses in which two or more parts are moved with respect to each other to obtain a continuously variable focal length and hence magnification, while the image is kept in the same image plane

If the diaphragm is opened at the same time so that its linear opening increases with the focal length it is possible to keep the relative aperture of the whole system constant while the focal length varies In any case, the system must be constructed so that the errors do not vary too much in the shifting Thus the designer must strive for a design in which the image errors are small, at least for the beginning and end positions of the "zooming" procedure and do not become too large for intermediate positions.

In general, a complicated cam is needed to control the motions of the parts of the system. How-



Zoomar lens in three operating positions. (a) Wide-angle. (b) Medium-angle (c) Telephoto Lens elements 2, 3, and 5 are mounted in movable barrel connected to zoom handle. Lens elements 1, 4, and 6 are stationary, *mounted in lens housing (From D. G. Fink, ed , Television Engineering Handbook, McGraw-Hill, 1957)*

ever, it is possible to simplify the mechanism if the focal plane is not kept precisely constant, but only required to coincide exactly with a given plane for several focal lengths while approximately coinciding for others Such a system is called an optically compensated varifocal system.

Some early variable-focal-length lenses contained 15 or more separated elements, but this number has been reduced to as few as 4 The zoom ratio (the ratio of maximum to minimum power) is in general 3:1, but it has been possible in some designs to increase the range to 4:1 or even more. Moving sets of achromatic prisms are sometimes used to achieve the zooming effect

The Zoomar lens, a variable-focal-length lens manufactured by the Television Zoomar Corporation, is shown in the illustration in three operating positions The focal length is determined by the television cameraman by means of a zoom handle which he pushes forward or pulls backward *See* LENS, OPTICAL [M.H.]

Bibliography. F Back, *J. Opt. Soc. Am* , 43(8): 685–689, 1953, L Bergstein, *J Opt Soc. Am*, 48(3) 154–171, 1958, H Chrétien, French Patent 766 526, B. Luboshez U S Patent 2,828,670, H H Hopkins, *Proceedings of the London Conference on Optical Instruments* (1950), 1951

Zoomastigophorea

A class of protozoans of the subphylum Mastigophora. Zoomastigophorea, also known as Zoomastigina, are flagellates which have few or no characters relating them to the pigmented forms Some are simple, some are specialized, some have pseudopodia besides flagella, others have no pseudopodia One group engulfs solid food at any body point, another shows localized ingestive areas. All are colorless None produce starch or paramylum, and lipids and glycogen are assimilation products Cells are naked or have delicate membranes Colony

rarely utilized independently. Most frequently the investigator weaves the several approaches together in an effort to determine the nature of a zoogeographic pattern, how it developed, and how it is sustained. [L.C.S.]

Bibliography: W. C. Allee and K. P. Schmidt, *Ecological Animal Geography*, 1951; D. Amadon, The Hawaiian honeycreeper (Aves, Drepaniidae), *Bull. Am. Museum Nat. Hist.*, 95(4), 1950; P. J. Darlington, *Zoogeography*, 1957; L. R. Dice, *The Biotic Provinces of North America*, 1943; E. R. Dunn, The herpetological fauna of the Americas, *Copeia*, 3:106–119, 1931; H. Gadow, *The Wanderings of Animals*, 1913; P. E. James and C. F. Jones (eds.), *American Geography, Inventory and Prospect*, 1954; W. D. Matthew, Climate and evolution, *Ann. N.Y. Acad. Sci*, 24:1–323, 1915; C. H. Merriam, Laws of temperature control of the geographic distribution of terrestrial animals and plants, *Natl. Geograph. Mag.*, 6:229–238, 1894; A. E. Ortmann, The geographical distribution of fresh-water decapods and its bearing upon ancient geography, *Proc. Am. Phil Soc.*, 41:267–400, 1902; P. L. Sclater, On the general geographical distribution of the members of the class Aves, *Proc. Linnaean Soc. London*, 130–145, 1858.

Zoological nomenclature

The system of names used to designate the various kinds of organisms which comprise the animal kingdom. First formalized by the eighteenth-century Swedish naturalist Linnaeus, the system has been expanded and modified into a universal procedure for the naming of animal species and their arrangement in a hierarchy of appropriately designated categories of classification (taxa). Until the adoption of this method, progress in the systematic study of animals was greatly hampered and the development of a natural classification scarcely possible.

Binomial nomenclature. In the Linnaean system of binomial nomenclature, the species has a dual designation, the first element of the name representing the genus (for example, *Canis*, the various dogs), the second, a particular species in that genus (as *lupus*, the wolf) These two elements comprise the species name, which must be unique; that is, the particular combination of generic and specific name cannot be used to designate any other species of animal It is followed by the name of the author. Thus, the correct scientific name for the wolf is *Canis lupus* Linnaeus. When, through inadvertance, or through reclassification two or more such combinations come into existance they become homonyms and the more recent are rejected. Likewise, each species can have only one valid name, the first properly proposed. Subsequent names become synonyms and are rejected (principle of priority). Since Greek and Latin were the common languages of educated men of the eighteenth century, it was natural to make use of these languages in forming scientific names, the generic names being selected largely from latinized Greek or from combinations of Greek words, the specific names mainly from Latin. Subsequently it has become customary to cite the name of the author following the binomial specific name, alone if the original combination of generic and specific names is still used, in parentheses if the combination has been changed. However, in order to be available for use a newly proposed name must be accompanied by a definition or description of the taxon to which it applies. A name proposed without such a definition is a nomen nudum, one with a definition which cannot be interpreted satisfactorily is a nomen dubium.

Higher taxa. Names of genera and other higher categories (taxa) in classification are single words (uninomial), and must be unique, not used for other groups of animals. The generic name is singular; those of higher categories are always plural Names for categories above the genus and below the order group are based upon the root of the name of one of the included genera (the type genus) and their ending indicates the category, as "-inae," subfamily; "-idae," family. Above the family group of names, designations for higher categories, as orders, classes and phyla, do not have uniform endings, although in practice there is considerable uniformity in certain groups such as in the Insecta where the ending "-ptera" usually denotes an order.

International code. Although Linnaeus proposed an elaborate series of rules governing the naming of organisms, in subsequent years these became variously modified in practice or altered to meet the requirements of expanding knowledge. As varying procedures were introduced into practice, it became imperative that steps be taken to secure a more stable and universal nomenclature if the basic system of Linnaeus was to survive. To this end, the International Congress of Zoology, toward the beginning of the twentieth century, established an *International Code of Zoological Nomenclature* (*Règles Internationales de la Nomenclature Zoologique*) and a permanent International Commission on Zoological Nomenclature to enforce and interpret it. This code consisted of 41 articles and 20 recommendations, dealing with family, generic, specific and subspecific names, their validity, formation, and orthography, the designation of types for species and genera and related matters. Of particular importance was the adoption of the principle of priority of publication, the separation of zoological nomenclature from botanical nomenclature and the selection of the tenth edition of Linnaeus' *Systema Naturae* (1758) as the starting point of zoological nomenclature This was an important international achievement However, the retroactive application of the rules to the zoological publications of nearly 150 years created many problems and jeopardized the stability of many names then widely utilized in the fields of medicine, public health, and agriculture As a result, the Commission was subsequently provided with plenary powers to set aside the application of the

development of a strong, disease-free livestock industry in these poverty stricken areas would provide animal protein to these starving people, improve their health status, increase their life span, help them to become more productive and so raise their standard of living

Use as a weapon. Some of the microorganisms causing zoonoses are adaptable for use in biologic *warfare* The use of these agents in time of war could wreak havoc upon the nations using them and upon noncombatant nations, as well as on the na-

and men, must establish and be ready to apply means of rapidly and accurately identifying the disease producing agent, and must be prepared to apply control measures

Spread of the zoonoses. Air travel, by reducing the transportation time, makes it possible for an animal or human in the incubation stage of a zoonosis to arrive in an area while still clinically well If the developing infectious disease is one with which the medical scientists in the receiving area are not familiar, an epizootic, epidemic, or both could occur.

Zoonoses control is further complicated by the insects that spread some of these diseases Man transports vectors from farm to farm community to community state to state, and nation to nation Some are carried with commodity shipments, others remain on the animals or in the straw and manure during transportation, and still others are carried in airplanes Changing environmental conditions also favor the establishment of different vectors.

The showing, or sale, of livestock at shows, fairs, or auctions further favors the spread of insect

Some bird species, sold as pets, are a potential source of disease Fortunately means of control have now been developed, unfortunately their application and acceptance will require time

While travel and movement of livestock and people are necessary they must be allowed to

World Health Organization Food and Agriculture Organization Expert Group on Zoonoses in its 1959 report lists five diseases due to infestation by insects The attacking insects are mites, bugs, fleas, flies, and ticks The principal host animals to these insects are poultry, birds, rodents, dogs cats, swine, equines and ruminants

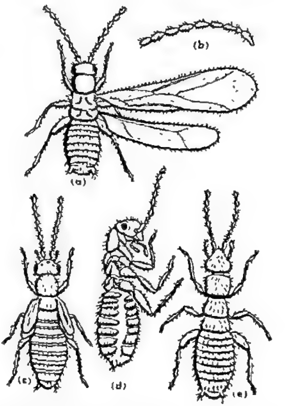
Conclusions. Medical scientists, human and veterinary, should expand their studies in comparative medicine The animal man environment relationship needs further elucidation More specific accurate, and faster methods of identifying and controlling infectious agents and the disease processes caused by them are needed. Scientists, of various disciplines, need to work more closely in studies of the epidemiology and epizootiology of the zoonoses and to apply the knowledge gained.

Early recognition and reporting to a central authority is essential to the institution of adequate control measures for each zoonosis if the weak link in the chain of infection is to be identified and broken. To accomplish this, people must be persuaded of the value of reporting to national centers which in turn feed to and receive information from an international epidemic and epizootic control center International agreements for control measures must be developed and utilized *See* ANIMAL VIRUS; BACTERIOLOGY, MEDICAL; ECOLOGY; PARASITOLOGY, MEDICAL. [J.D.M.]

Bibliography: R. W. Phillips, *Some Important* ... mal Diseases ... *ort of Joint* ... WHO Tech.

Zoraptera

An order of insects which are related to the termites and psocids They are about 2–2.5 mm long. Zoraptera are not important economically. They are of great interest because of their rarity, relatively



Zoraptera, *Zorotypus hubbardi* Caudell, greatest length 2.5 mm (a) Winged adult female (b) Antenna of wingless adult (c) Nymphal female of winged caste (d) Dealated adult female. (e) Adult female of wingless caste (From A N Caudell, *Proc. Entomol. Soc Wash., vol 22, 1920*)

formation is common. Colonies differ greatly in form and may be amorphous, linear, spherical, arboroid, or plane. Flagella vary from none to many. Nuclei generally have endosomes; frequently a rhizostyle extends from the nucleus to the flagellum base. Nutrition is holozoic, saprozoic, or parasitic, and encystment is common. Ecologically Zoomastigophorea are adapted to waters of relatively high organic content, but they occur in the clearest fresh waters and in the high seas. Four orders are usually included in this class: the Rhizomastigida, Protomastigida, Polymastigida, and Hypermastigida. See articles on these classes. *See also* MASTIGOPHORA. [J.B.L.]

Zoonoses

Animal diseases transmissible to man. The joint World Health Organization/Food and Agriculture Organization Expert Group on zoonoses in 1959 classified as zoonoses of major importance 94 of the more than 200 communicable diseases of animals. In addition five human diseases caused by insect infestations were classified as zoonoses. As investigation continues and investigative methods improve, the number of identifiable zoonoses will probably increase.

Personal, national, and international health and welfare are inextricably related to the health of animals. The zoonoses because of their epizootic and epidemic potentials are a constant threat and source of concern (*see* EPIDEMIOLOGY). If man is to maintain a biologic balance, to improve his health status, increase his longevity and productivity, and enhance his standard of living, he must adapt to, modify, control, or eradicate animal pathogens and parasites

Because of communal living and domestication of animals there has been a concentration of man and animals in a circumscribed area. This has resulted in a concentration of animal and human parasites and allowed greater exposure and disturbances in the host-parasite-environment relationship with consequent epizootics and epidemics. Anthrax abortions in cattle and man, bubonic plague, rabies, tsutsugamushi fever, and yellow fever are some of the zoonoses recorded in history. *See* ANTHRAX; PLAGUE; RABIES, TYPHUS, SCRUB; YELLOW FEVER

Present status. Animal diseases transferable to man are an unfinished chapter in history. Bubonic plague smolders in many endemic foci, if the right conditions occur it can again become epidemic or pandemic. African trypanosomiases not only affect man directly in their impact on his health, but they hold in economic poverty an area of otherwise fine grazing land in Central Africa, greater in size than the United States (*see* SLEEPING SICKNESS, AFRICAN). Mastitis is a problem in every country where dairy animals are used *Staphylococcus* mastitis is being recognized more frequently as a cause of food intoxications. Besides the diseases resulting from the drinking of raw milk, mastitis contributes to the protein starvation of the world. It is estimated that if mastitis were eliminated from the dairy herds of Europe alone milk production would be increased sufficiently to provide 1 pint each of milk a day for 30,000,000 children. *See* MASTITIS (COWS); STREPTOCOCCUS. Brucellosis in cattle causes even greater losses than does mastitis. *See* BRUCELLOSIS; STAPHYLOCOCCUS.

Zoonoses make occupational contact with infected animals a hazard. Diseases resulting from such contacts constitute a significant industrial health problem. Their number is great and each is a threat Anthrax, brucellosis, tuberculosis, Q fever, ornithosis, Newcastle disease, leptospirosis, and ringworm are just a few. *See* DERMATOPHYTOSIS; LEPTOSPIRA; NEWCASTLE DISEASE; PSITTACOSIS; Q FEVER.

As man's work economy has been developed to provide him with greater income and more leisure time, he has devoted more time to recreational activities. Many of these are affected by animals. Hunters are exposed to the threat of rabies, tularemia, yellow fever, and trypanosomiasis. Those whose recreation takes them into contact with birds expose themselves to the risk of ornithosis, psittacosis, and the arthropod-borne encephalitides. In those areas where the dog and cat hookworms are found, bathers and other users of beaches are plagued by creeping eruption Those who bathe or swim in stagnant waters expose themselves to leptospirosis. *See* HOOKWORM DISEASE; LOUPING-ILL; TICK FEVER, COLORADO; TUBERCULOSIS; WEST NILE FEVER.

As part of his social development man has made household and houseyard pets of many animals: dogs, birds, cats, rodents, and others In addition many wild animals have moved close to man's dwelling places because of the readily accessible supply of food As animals have been brought or moved into the immediate environs of man's living area, their parasites and pathogens and the vectors associated with them have been brought into close contact with man Pathogens, exotic to man, are now living in close proximity to him. Diseases caused by these pathogens are being recognized with increasing frequency in humans

On every continent there are vast animal reservoirs of pathogens capable of producing disease if the host-parasite-environment relationship be disturbed Man must learn to cope with these disturbances

Effect on food supply. In all of the world, except in a few areas, malnutrition is one of the major public health problems. About one-half of the people of the world are existing on diets deficient in animal protein. To overcome this need, which grows as the world's population increases, the world's technical knowledge, skills, and financial resources must be integrated in a continuing effort to control or eradicate the causes of animal diseases.

In those areas where zoonoses are prevalent at a high level and in which the livestock industry is underdeveloped, protein starvation is acute. The

recent discovery in 1913, and intriguing habits and relationships.

Zorapterons live sheltered from light, in decaying wood, usually in logs and stumps except during the emergence of winged adults. The adults are well pigmented, have eyes, and the wings are often shed. In the southeastern United States they occur especially on slabs buried in old sawdust piles. They apparently scavenge, mainly on microscopic molds. Most individuals are of the wingless caste, pale in color, and blind. Metamorphosis is gradual. The 9-segmented antennae and 2-segmented tarsi distinguish them from termites; the cerci separate them from psocids Mouthparts are for chewing.

Zoraptera are nearly world-wide in warm countries. The best known of the two United States species, *Zorotypus hubbardi*, occurs from Pennsylvania to Florida, west to Missouri and Texas. Approximately 21 species are known; all are in the genus *Zorotypus* of the family Zorotypidae. *See* INSECTA. [A.B.GU.]

Bibliography: A. B. Gurney, A synopsis of the order Zoraptera, with notes on the biology of *Zorotypus hubbardi* Caudell, *Proc. Entomol. Soc Wash.*, 40:57–87, 1938; P. P. Grassé (ed.), *Traité de Zoologie*, vol. 9, 1949.

Zythiaceae

A family of fungi of the order Sphaeropsidales, which contains many plant and insect pathogens. The family is also known as Nectrioidaceae. *Aschersonia* has been used in the biological control of scale insects in humid climates. There are 60 genera and 135 species. Many species are conidial stages of Hypocreales (order of Ascomycetes). Pycnidia, the fruit bodies containing conidia, are flask-shaped, conical, or lenslike; fleshy or waxy; brightly colored (yellow, orange, or red); and usually display a round pore.

Zythia is a genus with 1-celled hyaline spores (Hyalosporae). The pycnidia are separate and erumpent (bursting through the substratum of the host tissue); most species are stages of *Nectria*. *Z. resinae* is common in resin flux of conifers; *Z. fragariae* causes a leaf-blotch of strawberry.

Polystigmina is a genus with threadlike, curved conidia (Hyaloscolecosporae). The pycnidia are immersed in a reddish, foliicolous stroma (living on leaves). *P. rubra*, the imperfect stage of *Polystigma rubrum*, lives on the leaves of *Prunus*. *See* FUNGI IMPERFECTI; PLANT DISEASE; SPHAEROPSIDALES. [N.F.B.]